DOCUMENT RESUME

ED 463 388 UD 034 919

AUTHOR	Glewwe, Paul
TITLE	Schools, Skills and Economic Development: Education Policies, Student Learning and Socioeconomic Outcomes in Developing Countries. Bulletin.
INSTITUTION	Minnesota Univ., Minneapolis. Dept. of Economics.; Minnesota Univ., St. Paul. Dept. of Applied Economics.
REPORT NO	Bull-01-3
PUB DATE	2001-12-00
NOTE	76p.; Produced by the Economic Development Center.
AVAILABLE FROM	For full text: http://agecon.lib.umn.edu/cgi-bin/pdf_view.pl?paperid=3839&ftype=.pdf.
PUB TYPE	Information Analyses (070)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Cognitive Development; *Developing Nations; *Economic Development; Educational Policy; Elementary Secondary Education; Foreign Countries; Labor Market; Learning; *Outcomes of Education; Productivity; *Socioeconomic Influences

ABSTRACT

This paper reviews recent research on the determinants of educational outcomes and the impact of those outcomes on other socioeconomic phenomena. It investigates the relationship between education and economic growth and development in emerging countries. The paper addresses school policies that are most cost-effective in producing students with particular cognitive skills, such as literacy and numeracy; the relationship between schooling, particularly cognitive skills acquired in school, and labor productivity; and the impact of schooling, especially cognitive skills, on other socioeconomic outcomes. Until recently, most empirical studies that examined what school policies were most cost-effective in producing students with particular cognitive skills estimated production functions for cognitive skills. This approach has serious shortcomings, which has led to new approaches such as generating or finding random variation in school characteristics and comparing cognitive skills across schools with different levels of these characteristics. Research shows that cognitive skills are directly responsible for part, if not most, of the impact of schooling in labor income. There is no evidence that innate ability has a direct impact on labor income. There is very little evidence on the impact of schooling on socioeconomic outcomes other than labor productivity. Recommendations for future research are presented. (Contains 81 references and 28 footnotes.) (SM)

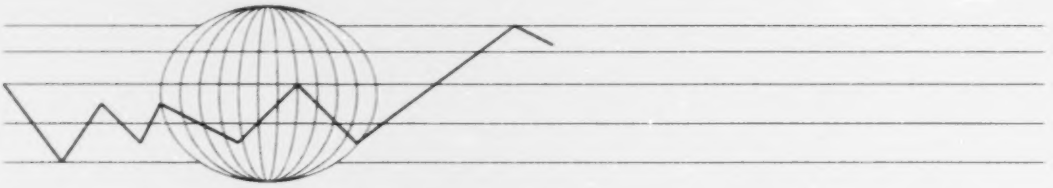
Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 463 388

Bulletin Number 01-3

December 2001

ECONOMIC DEVELOPMENT CENTER



SCHOOLS, SKILLS AND ECONOMIC DEVELOPMENT: EDUCATION POLICIES, STUDENT LEARNING AND SOCIOECONOMIC OUTCOMES IN DEVELOPING COUNTRIES

PAUL GLEWWE

ECONOMIC DEVELOPMENT CENTER
Department of Economics, Minneapolis
Department of Applied Economics, St. Paul

UNIVERSITY OF MINNESOTA

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

L Letnes
Economic Development Center, University of Minnesota

UD 034 919

BEST COPY AVAILABLE

Schools, Skills and Economic Development: Education Policies, Student Learning and Socioeconomic Outcomes in Developing Countries[1]

(forthcoming, *Journal of Economic Literature*)

Paul Glewwe
University of Minnesota and the World Bank

November, 2001

Abstract

This paper reviews recent research on the determinants of educational outcomes, and the impact of those outcomes on other socioeconomic phenomena. More specifically, it addresses three questions: 1. What school policies are most cost-effective in producing students with particular cognitive skills, such as literacy and numeracy? 2. What is the relationship between schooling, particularly cognitive skills acquired in school, and labor productivity? 3. What impact does schooling, especially cognitive skills, have on other socioeconomic outcomes? While recent research has made some progress, these are difficult questions and much more work is needed. The paper provides suggestions for future research on these questions.

[1] I would like to thank the following people for comments, discussions and/or clarification on their papers: Bruce Fuller, Nancy Gillespie, Eric Hanushek, Emmanuel Jimenez, Dean Jolliffe, Cigdem Kagitcibasi Geeta Kingdon, Michael Kremer, Julia Lane, Berk Ozler, Lant Pritchett and Jee-Peng Tan. I am also grateful to John McMillan and three anonymous referees for very detailed and useful comments. The findings, interpretations, and conclusion expressed in this paper are entirely those of the author. They do not necessarily represent the views of the World Bank.

Introduction

Economists have studied economic growth and development since Adam Smith set out to explain the nature and causes of the wealth of nations. In the 1950s and 1960s, Gary Becker, Jacob Mincer, T.W. Schultz and others turned economists' attention to education and the role it plays in a variety of economic phenomena. More recently, economists have linked these two literatures, examining the impact that education can have, and in some countries already has had, on economic growth (Robert Lucas, 1988; Robert Barro, 1991; N. Gregory Mankiw, David Romer and David Weil, 1992). While none of these more recent studies is beyond criticism, few economists would claim that education has little or no role to play in promoting economic growth and development in low- and middle-income countries.

The proposition that higher levels of education promotes economic growth and development suggests that governments in developing countries should implement policies that raise educational attainment, since growth and development are objectives of nearly all developing countries. Thus many economists and international organizations argue that investments in education are a policy priority (Becker, 1995; Eric Hanushek, 1995; UNDP, 1990; World Bank, 2001). Yet at this crucial point economists often have no further recommendations to offer. That is, until very recently they have said little about what governments in developing countries can do to raise educational attainment.

This lack of advice does not imply that schools in developing nations are already operating effectively and efficiently. To the contrary, there is ample evidence that many schools in these countries are not very effective and as such operate far from any conceivable efficient frontier (Marlaine Lockheed and Adriaan Verspoor, 1991; Ralph Harbison and Hanushek, 1992; Hanushek, 1995; Paul Glewwe, 1999a). It is also not the case that governments and schools know how to improve educational outcomes but choose not to do so because such actions would not be in their interest. While there are situations where teachers and other education officials favor their interests over those of students, it is also clear that Ministries of Education in developing countries often are not sure what to do to improve their education systems (Lockheed and Verspoor, 1991, p.39).

This unsatisfactory state of affairs is all the more glaring given the staggering amounts of money involved: each year the governments of developing countries spend about $260 billion on education.[2]

Finally, this lack of knowledge on how to operate schools most effectively does not reflect lack of interest on the part of researchers. Many studies have addressed these issues, but most of them suffer from serious shortcomings. Recently, more careful studies have provided more reliable findings on some specific policies and programs. The purpose of this paper is to examine this recent work in detail.

More specifically, this paper has three objectives. The first is to review the literature on the relationship between school and teacher characteristics, broadly defined, and the acquisition of cognitive skills. The question addressed is: What school policies are most cost-effective in producing students with particular cognitive skills, such as literacy and numeracy? The second objective is to examine the relationship between schooling and labor productivity, with emphasis on the relationship between basic cognitive skills and labor productivity. Knowledge of the impact of different skills on income and on other socioeconomic outcomes could have policy implications for school curriculum. For example, if literacy were identified as more important than, say, scientific knowledge in determining future income, it may be desirable to reduce classroom time devoted to science in order to increase the time devoted to language skills. The third objective is to investigate the relationship between cognitive skills and socio-economic outcomes other than labor productivity, such as the impact of schooling on women's fertility and on adult and child health. The three main sections of this paper cover each of these objectives in turn. A final section summarizes the findings and provides recommendations for future research.

Before proceeding, a few comments are needed on the scope of the paper. First, it does not address the issue of whether government subsidies for education can be justified in terms of standard economic theory. Other papers have argued that this is the case (see,

---

[2] This figure is calculated by taking the total GNP of low and middle income countries in 1999, which amounted to $6,311 billion, and multiplying it by the (average) government

inter alia, Daron Acemoglu, 1996, and Roland Benabou, 1996), and this paper need not take a position on this issue. Second, while the paper considers the issue of whether private schools are more efficient than public schools, it also considers, in detail, what governments can do to improve the operation of public schools even though private schools may be more efficient. The reason for this is simple realism – many governments favor public schools for a variety of "non-economic" reasons (examples are perceived equity benefits and political objectives such as promoting a curriculum that gives students a national, as opposed to an ethnic or regional, identity) and thus policy advisors have little choice but to accept this constraint and focus on ways to improve public schools.

A final limit on the scope of this paper concerns the educational outcomes examined. Schooling provides children with many benefits. The most obvious are cognitive skills such as literacy, numeracy, scientific knowledge, and advanced thinking skills. Schooling can also provide social skills and (internalized) values that may help children succeed in the adult world. Lastly, prestige may be attached to particular levels of education, which may enable one to find a better job or a "better" spouse. A thorough study of all these benefits could double the length of this paper. To keep the paper to a reasonable length, it focuses on the basic cognitive skills that school curricula are designed to impart. However, occasional reference is made to other benefits of schooling.

I. School Characteristics and the Acquisition of Cognitive Skills

This paper approaches education issues from an economic perspective. That is, it takes the position that a model of "rational" behavior is needed to ensure that proper econometric and statistical methods are used to estimate the impact of school characteristics and policies on educational outcomes, and of the impact of schooling and cognitive skills on socioeconomic outcomes. In particular, explicit models of human behavior provide substantial insight into whether assumptions underlying specific econometric methods are satisfied. If a plausible model suggests that some assumptions

---

expenditures on education as a percentage of GNP, which was 4.1%. Both of these figures are from World Bank (2001).

are not satisfied, empirical findings based on those methods may be invalid. The model may also suggest how to test the econometric assumptions, and what estimation method can be used if they fail to hold. The section first presents such a model and examines its implications for empirical analysis. The model is not intended to be the definitive model of schooling, rather it is a simple yet plausible model that illuminates several econometric issues. After presenting the model and its implications for empirical work, several recent studies of the impact of school and teacher characteristics on learning are examined.

A. A simple model of schooling choices

Assume that parents make decisions for their children and that their objective is to maximize a utility function that has two arguments: consumption of goods and services and child cognitive skills. For simplicity, assume that there are two time periods and only one child per family.[3] In period 1, a child may attend school, work, or both. If both, the child first goes to school, working only after his or her schooling is completed (going to school first is optimal in most cases, see Chapter 3 of Glewwe, 1999a). In period 2, the child becomes an adult and works. When a child works in either time period, part or all of the child's earnings may be given to his or her parents. A utility function that takes parents' consumption (C) in periods 1 and 2 and child cognitive skills (A) as its arguments is:

$$U = C_1 + \delta C_2 + \sigma A, \qquad (1)$$

where δ is a discount factor for future consumption and σ indicates parental tastes for educated children (higher values imply greater utility from educated children). Parents value educated children for two distinct reasons: educating children can increase parents' consumption, and educating children directly affects parents' utility (through σ).

A simple production function shows how cognitive skills, A, are acquired:

[3] This implies that the number of children a family has is exogenous. Glewwe (1999b) considers the possibility that the number of children is a choice variable. That paper develops in more detail the model presented here.

$$A = \alpha f(Q)g(S), \tag{2}$$

where α is the "learning efficiency" of the child, Q is school quality, and S is years of schooling. The functions f( ) and g( ) are increasing in Q and S, respectively. A child's learning efficiency, α, represents several different factors, such as innate (genetically inherited) ability, child motivation, and parental motivation and capacity to help children with their schoolwork. For simplicity, all these factors are combined into α.

Parents' consumption in each time period is given by:

$$C_1 = Y_1 - pS + (1 - S)kY_c \tag{3}$$

$$C_2 = Y_2 + kY_c \tag{4}$$

where p is the price of schooling,[4] Y_1 and Y_2 are parental income in periods 1 and 2, respectively, Y_c is the child's income when working, and k is the fraction of that income given to the parents. The last term in (3), $(1 - S)kY_c$, requires some explanation. S has been rescaled to be the fraction of time spent in school by the child in time period 1. The remaining time in the first period, 1 - S, is spent working. This is purely for notational convenience; however, to keep the vocabulary simple, S is still called "years of schooling."

Equations (3) and (4) rule out borrowing and saving; the only way to transfer income between periods 1 and 2 is to alter investments in children's education. This assumption is made for simplicity. In general, introducing borrowing and saving would reduce parents' incentive to invest in their children's education. Yet it would not completely eliminate this incentive because almost all investments are risky, so most parents would diversify their investments among several different alternatives, including their children's education.

Equation (5) completes the model, relating child cognitive skills to child income:

---

[4] The child's consumption while in school can be included in p, while the child finances his or her own consumption from Y_c when working. Strictly speaking, this assumes that child consumption while in school is exogenous, perhaps set by local cultural norms.

$$Y_2 = \pi A, \qquad (5)$$

where π is the productivity of cognitive skills in the labor market.

Substitution of (2) into (5), of (5) into (3) and (4), and of (2) - (4) into (1), expresses parents' utility as a function of years of schooling (S) and school quality (Q):

$$U = Y_1 - pS + \delta Y_2 + ((1 - S + \delta)k\pi + \sigma)\alpha f(Q)g(S) \qquad (6)$$

Consider first the case where school quality is exogenous, so that S is the only choice variable. It is straightforward to derive the impacts of changes in the model's various parameters on the optimal (utility-maximizing) value of years of schooling (see Glewwe, 1999b), all of which are intuitively plausible. Optimal years of schooling (and thus the child's cognitive skills) is an increasing function of: the child's learning efficiency (α), school quality (Q), the relative weight (δ) parents give to future consumption, and parental tastes for schooling (σ). Optimal years of schooling decreases when the price of schooling (p) rises. Finally, optimal years of schooling is likely, though not certain, to rise when parents expect to receive a larger proportion (k) of their children's income from working and when the value of cognitive skills in the labor market (π) is higher. The intuition for this ambiguity is that although a higher value of cognitive skills in the labor market (π) raises the value of schooling, it also makes time out of school (which increases when years of schooling declines) more valuable. The same argument applies to the proportion of children's income going to parents (k).

The model is easily extended to allow parents to choose school quality (Q). Assume that parents choose school quality, but higher quality implies a higher price:

$$p = p_0 Q \qquad (7)$$

where p_0 is the "base" price of schooling. While (7) may appear to impose an arbitrary linear functional form (why should the price double if quality doubles?), this is not the case. One should interpret Q as an index of expenditures on quality. Whether, say, doubling expenditures doubles the impact of school quality on learning (that is, doubles f(Q)) depends on the functional form of f( ).

Replacing p with p_0Q in (6) yields an expression to be maximized with respect to S and Q:

$$U = Y_1 - p_0QS + \delta Y_2 + ((1 - S+\delta)k\pi + \sigma)\alpha f(Q)g(S) \qquad (8)$$

To simplify derivation of the impacts of changes in the various parameters on (optimal values of) S and Q, one more assumption is needed on the functional forms of f( ) and g( ). A convenient and plausible assumption is that $f(Q) = Q^{\beta}$ and $g(S) = S^{\gamma}$. Different values of β and γ yield a wide range of the shapes for both functions. Both β and γ must be positive to ensure that f( ) and g( ) are increasing in Q and S, respectively. While this assumption implies that the following results are not completely general, the model is still useful because it demonstrates the implications of plausible assumptions for empirical analysis.

Using these functional form assumptions, one can show (see Glewwe, 1999b) that the optimal values (denoted by asterisks) of S and Q are:[5]

$$S^* = (\gamma - \beta)(1 + \delta + \sigma/k\pi)/(1 + \gamma - \beta) \qquad (9)$$

$$Q^* = (\alpha\beta k\pi/p_0)^{1/(\gamma - \beta)}((1 + \delta + \sigma/k\pi)/(1 + \gamma - \beta))^{\gamma/(\gamma - \beta)}. \qquad (10)$$

The optimal level of cognitive skills (A) is obtained by inserting (9) and (10) into (2).

---

[5] Note that $S^* > 0$ only if $\gamma > \beta$. Intuitively, if $\gamma < \beta$ then cognitive skills (A, which equals $\alpha Q^{\beta}S^{\gamma}$) could be increased by doubling Q while halving S without any increase in the cost of schooling (which is p_0QS). This implies that S should approach zero while Q should approach infinity. Requiring γ to exceed β rules out such a corner solution.

These optimal levels of years of schooling (S) and school quality (Q) are intuitively plausible. Both increase when parents put more weight (δ) on future consumption and when parents have higher tastes for schooling (σ). School quality (Q) increases with learning efficiency (α) but decreases as the base price of schooling (p_o) rises. A less plausible result is that years of schooling depends neither on the base price of schooling nor on learning efficiency. This reflects the functional forms of f( ) and g( ), but is not necessarily unreasonable. Basically, when the base price falls or child learning efficiency rises, parents shift to higher school quality, raising their children's cognitive skills without changing years of schooling. By choosing higher quality instead of more time in school, parents avoid a cost of the latter: reduced child working time in period 1 (see equation (3)). In developing countries, grade repetition is high, so this can take the form of reduced grade repetition, raising the highest grade attained without changing years of schooling.

A second apparently counterintuitive result is that increases in the propensity of children to support their parents (k) and in the market return to cognitive skills (π) decrease years of schooling. Yet these results may be reasonable: one response to such changes is to choose higher school quality and reduce time spent in school to increase the time the child spends working in time period 1.[7] Of course, other functional forms for f( ) and g( ) could lead to different impacts of k and π on years of schooling.

This simple model produces many intuitively plausible results. It also provides some insights that go beyond simple intuition. For example, when school quality is exogenous it is not necessarily intuitive that parents who give greater weight to future consumption will send their children to school longer, even after controlling for parental tastes for schooling. Even less obvious is the result that higher returns to cognitive skills do not necessarily increase years of schooling (because they raise the opportunity cost as well as the benefit of an additional year of school). When school quality is also a choice variable, the main insights beyond simple intuition work through the fact that years of schooling and school quality are alternative inputs in the production of cognitive skills. This explains why the

(base) price of schooling has no effect on time in school; the best response to a change in this price may be to adjust school quality, holding years of schooling constant (although the highest grade attained may rise due to less grade repetition). While the absence of any effect on years of schooling reflects functional form assumptions, under almost any functional forms one should find that the impact of the price of schooling on years in school diminishes when school quality becomes endogenous. Similarly, the increase in years of schooling due to an increase in a child's learning ability is smaller when parents have the option of increasing school quality. A final insight from this model when school quality is endogenous is that the price of schooling *per year* of enrollment at the chosen school, p_iQ, is an endogenous variable; econometric analyses should not treat school prices at the school attended as exogenous.

B. Implications of the model for econometric analysis

The model presented above provides a useful framework for discussing several issues concerning estimation of the impact of school characteristics on cognitive skills. Most empirical studies that attempt to estimate the cognitive skills production function given in (2) assume linear functional forms to simplify estimation.[7] Thus (2) becomes:

$$A = \mu_0 + \mu_1 S + \mu_2 \alpha + \mu_3 Q + e \qquad (2')$$

where the μ coefficients are unknown parameters to be estimated. The simplest interpretation of the residual term e is that it reflects measurement error in A, but of course it could reflect omitted variables, or measurement error pertaining to α, Q and even S.

---

[6] School quality will most likely rise when k or π increases, but it could decrease. The intuition for a decrease is that although cognitive skills must decline, this loss in parental income is outweighed by the increased income from the child working longer in period 1.

[7] Linearity can follow from the model presented above. Taking the logarithm of both sides of (2) and assuming exponential functional forms for f( ) and g( ), such as $f(Q) = Q^{\beta}$ and $g(S) = S^{\gamma}$ yields an equation that is linear in the logarithms of the variables.

The specification of school quality in (2') is clearly oversimplified. It is more realistic, and more useful for policy analysis, to decompose school quality into a function or index of the different school characteristics that promote learning:[8]

$$A = \mu_0 + \mu_1 S + \mu_2 \alpha + \tau_1 Q_1 + \tau_2 Q_2 + ... + \tau_n Q_n + e. \qquad (2'')$$

In (2''), Q is replaced by an index of n distinct school characteristics that affect learning. Policymakers would like to know the magnitude of the various τ's because such estimates can be combined with data on the costs of those same school characteristics to assess the cost-effectiveness of each characteristic in promoting learning. Indeed, this information is precisely what is needed to answer the first of the three questions addressed by this paper, namely which school policies are most cost-effective for raising student's cognitive skills.

A child's learning efficiency, α, is also multi-dimensional. Some factors that raise learning efficiency, such as parental education, are easily observed, while collecting data on others is very difficult, if not impossible. Thus (2'') can be rewritten as:

$$A = \mu_0 + \mu_1 S + \rho_1 \alpha_1 + \rho_2 \alpha_2 + \ldots \rho_m \alpha_m + \tau_1 Q_1 + \tau_2 Q_2 + ... + \tau_n Q_n + u. \qquad (2''')$$

In this equation the *observed* components of learning efficiency are specified as α_1, α_2, etc. In contrast, the *unobserved* components must be combined with e, which yields u, a residual term that represents both random measurement error in A and the impact of unobserved aspects of learning efficiency (α) on cognitive skill acquisition (A). In fact, u also represents unobserved school quality characteristics, as well as measurement error in S and in the α and Q variables.

Examples of difficult to observe learning efficiency variables are the child's innate ability and motivation, and parents' willingness and capacity to help their children with schoolwork. One can try to measure some of these factors (such as using an IQ test to

[8] This linear function of the school characteristics can be made more realistic by adding quadratic and interaction terms. To simplify the exposition, these terms are omitted.

measure innate ability and using parental schooling to indicate parents' ability to assist their children), but it is unlikely that one can measure all of them. Indeed, it is not clear that innate ability can be measured: any test that claims to do so (in the sense of measuring a genetic endowment) almost always reflects environmental factors (American Psychological Association, 1995). One may be able to avoid this problem by using data on twins (for example, Jere Behrman et al., 1994), but such data from developing countries are very rare.

Many aspects of school quality are also unobserved. Most data sets have only a small number of school quality variables; many easy to observe school characteristics are often omitted when the data are collected. In addition, some aspects of school quality are inherently difficult to measure, such as teachers' interpersonal skills and motivation, and the management skills of school principals.

Suppose that (2''') is estimated using ordinary least squares (OLS). Of course, the estimated parameters are unbiased only if the residual, u, is uncorrelated with S and the various Q's and α's. Yet the model presented in the previous subsection shows that such correlation is very likely: in equation (10), higher learning efficiency (α) increases school quality (Q), implying that u, which contains the unobserved components of α, is positively correlated with the various Q's. Thus estimates of the associated parameters (τ's) will be biased upwards. The estimated impacts of observed learning efficiency variables are also likely to be biased, since those variables are usually correlated with unobserved aspects of learning efficiency. Most empirical studies do little or nothing to avoid this problem.

If school quality were exogenous, one might think that these estimation problems can be avoided because coefficients on any exogenous variables are unlikely to be biased. Yet econometric theory shows that correlation between *any* variable and the error term is likely to lead to biased estimates of *all* parameters, not just the parameter of variables with which the error is correlated (Russell Davidson and James MacKinnon, 1993, pp.211-215). In the simple model given above, years of schooling is positively correlated with learning ability when school quality is exogenous, which will lead to biased estimates for the school quality parameters.

Moreover, school quality is likely to be endogenous. Even in rural areas of low income countries, where villages often have only one school and are too far apart for children to attend school in a neighboring village, parents may be able to influence school quality. First, they may directly alter the quality of the sole local school through the parent-teacher association (PTA) or through political connections. Second, they may send their children to live with relatives (allowing them to attend a nonlocal school) or to a boarding school. About 19 percent of secondary students in rural Peru live away from their families (Paul Gertler and Glewwe, 1990), and the same holds for 27 percent of middle school students in Ghana (Glewwe and Hanan Jacoby, 1994). Third, families with higher tastes for educated children may migrate to areas with better schools, a common occurrence in the U.S.

When parents can alter school quality, overestimation is possible due to positive correlation between unobserved components of a child's learning efficiency and school quality. Endogenous school quality can also lead to underestimation. Even when parents cannot alter school quality, quality could be correlated with the error term if governments provide better schools to areas with unobserved education problems (Mark Pitt et al., 1993). These unmeasured problems would also be relegated to u in equation (2'''), producing negative correlation between the error term and the school quality variables (Q's) and thus underestimating the impact of school quality. On the other hand, governments are just as likely (and some would argue much more likely) to place better schools in areas that already have good education outcomes since both autocratic and democratic rulers often derive political support from elite groups (World Bank, 2001). For empirical evidence on this point, see Nancy Birdsall (1988) for Brazil and Behrman and James Knowles (1999) for Vietnam.

In theory, instrumental variable methods can resolve this problem, but it is difficult to find plausible instruments. One possible instrument for years of schooling is the price of schooling, which should affect learning only by affecting years of schooling. Alternatively, one could estimate (2''') for a single grade to remove variation in S. Yet both approaches have problems. First, the prices observed in the data for the schools attended are not the p_0's of equation (7) but p_0Q, which is endogenous if Q is endogenous. In particular, it will

be correlated with u, invalidating its use as an instrument. Second, if some children in the relevant age range are not in school, the remaining children (whether in one or several grades) are not a random sample of the population. Intuitively, communities with high-quality schools will keep children in school longer, leading to a student population with lower average learning efficiency (more "less-efficient" children stay in school). In this case u in (2''') will be negatively correlated with school quality, leading to underestimation of the impact of school quality on learning. Third, no data set includes every component of school quality, and observed components may be positively correlated with unobserved components (because "good" schools are often good in many ways, only some of which are observed). Again, unobserved aspects of school quality are part of the residual in (2'''), causing u to be positively correlated with observed school quality variables and causing the τ parameters to be overestimated.

A final difficulty in empirical work is measurement error in the explanatory variables, both S and the various Q variables. Random measurement error will cause underestimation of the impact of both S and Q on the acquisition of skills, while nonrandom measurement error could lead to underestimation or overestimation.

In summary, uncritical application of simple OLS regressions can lead to biased estimates of the impact of school quality on learning. Some problems underestimate the impacts, others overestimate them, and still others could go either way. These difficulties are so daunting that some economists doubt that they can be overcome (see the pessimistic assessment of Hanushek, 1995). The next two subsections examine several recent studies, focusing on how these problems have been addressed, or not addressed, in the literature.

C. Recent Estimates of the Impact of School Characteristics on Student Skills

How have studies of the impact of school characteristics on students' cognitive skills dealt with the problems raised above? More generally, how much has been learned that governments can apply to make schools more effective? This subsection reviews "conventional" studies by education specialists and economists, where conventional refers to studies that attempt to estimate educational production functions along the lines of

equation (2''') using ordinary (non-experimental) variation in the explanatory variables. The following subsection examines several more recent, and more innovative, papers.

Most conventional studies of the impact of school characteristics on learning focus on developed countries, yet research on developing countries has increased rapidly in recent years. Bruce Fuller and Prema Clark (1994) provide a comprehensive review of the literature up through the mid 1990s. Earlier literature reviews can be found in Harbison and Hanushek (1992) and Fuller (1987). While these reviews are comprehensive, they tend to take the conclusions of the studies they review at face value. Many economists who have examined these studies find serious methodological shortcomings. For example, Hanushek (1995, pp.231-232) claims that "...the standards of data collection and analysis are so variable that the results from this work are subject to considerable uncertainty." Anne Case and Angus Deaton (1999, p.1081) concur, stating that "the descriptions of econometric procedures...are sometimes so exotic as to raise serious doubts about the validity of the results." My own reading of the conventional literature confirms that the estimation methods used typically ignore most of the problems raised in the previous subsection.

Given these methodological shortcomings, it is not surprising that the findings of some studies are at odds with those of others. Fuller and Clark's summary conveys the uncertainty in the literature regarding key questions. While many observers would expect reductions in class size to increase learning, Fuller and Clark find that only 9 of 26 primary school studies and only 2 of 22 secondary school studies show a significant impact of class size on student achievement in developing countries. Moreover, the paper reports only significant effects that are in the expected direction (for example, smaller class size raises educational achievement).[4] Ignoring significant effects in unexpected directions may be misleading: the literature summary by Harbison and Hanushek (1992) included 30 studies that examined the impact of teacher-pupil ratios and found that of the 16 with statistically significant effects eight were positive and eight were negative! These problems cast doubt on whether any conclusions can be drawn with confidence from the conventional literature.

[4] I would like to thank Bruce Fuller for explaining this to me.

This pessimistic interpretation includes meta-analyses along the lines suggested by Michael Kremer (1995), since that approach is only as plausible as the studies on which it is based.

Can more careful conventional estimates produce useful results? The rest of this subsection addresses this question. Before doing so, an important point needs to be made regarding studies that have attempted to draw inferences about school quality based on wage data, such as David Card and Alan Krueger's (1992) study of U.S. schools and Behrman and Birsdall's (1983) study of Brazil. The point is that very little can be inferred from such studies regarding what makes one school better than another, because such studies typically have only one indicator of school quality, such as spending per pupil or the average education level of teachers. Clearly, any single indicator of school quality is likely to be correlated with many other school quality variables, so such studies cannot determine which school variables improve children's learning. To make further progress, data are needed on schools, teachers and students' cognitive skills.

Four studies completed in the early to mid 1990s attempted to estimate educational production functions using data specifically collected for that purpose: Harbison and Hanushek's (1992) book on Brazil; Glewwe and Jacoby's (1994) study of Ghana; the analysis Jamaican data by Glewwe et al (1995); and Geeta Kingdon's (1996a) paper on India. These are probably the best "conventional" studies, so it is worthwhile to see how they address, or do not address, the problems raised in subsection I.B and, more generally, how useful their results are for making education policy decisions in developing countries.[10]

Harbison and Hanushek examined the performance of primary school children in rural areas of Northeast Brazil in reading (Portuguese) and mathematics. Tests were administered in 1981, 1983 and 1985. The school characteristics examined were a facilities index (of about 10 building characteristics), a writing materials index (chalk, notebooks, pencils, etc.), the availability of textbooks, and a dummy variable indicating graded classrooms (as opposed to multi-grade classrooms). Both the facilities and the writing

[10] The study of Pakistan by Harold Alderman et al (1996a) is not discussed here because it does not estimate the impact of school or teacher characteristics on student achievement.

materials indices had significantly positive impacts in most specifications for both reading and math. The textbook variable was significantly positive for three of five specifications in math and two of five in reading. Graded classrooms was never significantly positive; in some cases it was significantly negative. The study also examined teacher characteristics. Neither the pupil-teacher ratio nor teacher experience had consistent effects in either subject, but teacher salaries had significantly positive impacts in both subjects. Teacher education almost always had insignificant impacts for reading, but usually had a significantly positive impact for math. Finally, the impact of teacher training programs was mostly insignificant.[11] To give an idea of the size of one of the significant impacts, consider the teacher salary variable. In the 1983 level specification (second grade students), doubling teachers' salaries raised reading test scores by 0.14 standard deviations and math scores by 0.15 standard deviations. These effects are not particularly large compared to those of the three other studies, as will be seen below.

In Glewwe and Jacoby's study of Ghana, achievement tests were given in 1988-89 in reading (English) and mathematics in middle schools (grades 7-10). Many school and teacher variables were examined. Most estimated effects were small and not statistically significant. The only statistically significant teacher variable was teaching experience, but its effect was only indirect; it raised children's grade attainment, which then increased both reading and math test scores. The estimated impact of repairing leaking classrooms, which presumably reduced school closings due to rain, was much larger; the overall (direct plus indirect) impact was an increase of 2.0 standard deviations in reading scores and 2.2 in math scores. Blackboards also had large estimated impacts (direct plus indirect), raising reading scores by 1.9 standard deviations and math scores by 1.8. Adding a library led to smaller increases, 0.3 standard deviations for reading and 1.2 for math scores.

The Jamaica study used data collected in 1990 on the performance of primary school students in reading (English) and mathematics. Over 40 school and teacher characteristics were examined, including pedagogical processes and management structure. Most variables had statistically insignificant effects. The school variables with significantly

[11] Another explanatory variable in the Brazil study is teachers' test scores, but the table and discussion in the text (p.114) contradict the results given in the appendix tables.

positive impacts were administration of eye examinations (reading only), teacher training within the past three years (math), routine academic testing of students (reading and math), and the use of textbooks in class (reading). Class time devoted to written assignments had a significantly negative impact in both subjects. The size of these estimated impacts (in standard deviations of the test score variable) were lower than those for Ghana. The largest impact is a change from never using textbooks in instruction to using them almost every lesson, which raises reading scores by 1.6 standard deviations. The smallest is from teacher training: a school in which all teachers were trained is estimated to have math scores 0.7 standard deviations higher than an otherwise identical school with untrained teachers.

Kingdon's study of India is based on data collected in 1991. Tests in reading (Hindi and English) and mathematics were given to students in "class 8" (grade 8). She examined five teacher variables (years of general education, years of teacher training, marks received on official teacher exams, years teaching experience, and salary) and three school variables (class size, an index of 17 physical characteristics, and hours per week of academic instruction). The teacher variables with significant effects were teacher exam marks, which had significantly positive impacts on both math and reading scores, and teachers' years of education, which had a significantly positive impact on reading scores. Two of the three school variables, the physical characteristics index and time in academic instruction, had significantly positive effects on both reading and math scores. Class size has no significant impact on math, and a significantly *positive* impact on reading. The impact of the teacher's exam marks is not robust to attempts to control for selection into schools (an issue discussed further below). These impacts are not particularly large. An additional year of teacher's education raises reading scores by 0.13 standard deviations. Going from zero to all 17 physical facilities (which would be quite costly since this includes toilets, computers and musical instruments) increases math scores by 0.7 standard deviations and reading scores by 1.0 standard deviations. Adding another hour per week of instructional time raises math and reading scores by only 0.04 and 0.02 standard deviations, respectively.

How much confidence can be placed in the results of these studies? Of the issues raised in the previous subsection, consider first the problem that unobserved components of a child's learning ability, such as a child's innate ability and motivation and parent's

willingness to help their children with their schoolwork. This leads to upwardly biased estimates of the impact of school quality variables. The Ghana and India studies used data from an "intelligence" test, the Raven's Coloured Progressive Matrices test, to control for innate ability. The Ghana study concedes that this test measures not only innate ability (however defined) but also reflects environmental influences, including time in school. It used a simple "family fixed effects" procedure to extract what is probably a cleaner estimate of innate ability from the Raven's test, but this method relies on several rather simplistic assumptions. The India study used the Raven's test score directly, without any refinement, and the Brazil and Jamaica studies had no variables to control for child innate ability. Only one of the four studies, the one on India, attempted to control for child motivation as a factor that is distinct from innate ability. (Another possible exception is the value-added estimates in the Brazil study, which are discussed below.) Regarding parents' motivation and ability to help their children, none of these studies goes beyond the common practice of using mother's and father's years of education. On a more positive note, all of these studies use standard selectivity correction methods (primarily to account for choices among different types of schools); this may reduce bias caused by a variety of unobserved variables, including innate ability.

Another potential problem is bias due to omitted school and teacher quality variables. If unobserved school and teacher variables are positively correlated with observed school and teacher variables, the estimated impacts on the observed variables will be biased upwards. At first glance, all four studies seem to minimize this problem by including large numbers of school and teacher variables. The Brazil study used data on at least 20 school and teacher characteristics (the exact number is unclear because many were aggregated into indices). The original Ghana study used 18 school variables (see Glewwe and Jacoby, 1992), and the Jamaica study had 42, including variables on pedagogical techniques and "school organization, climate and control". Finally, the India study used data on about 24 variables, although 17 were aggregated into a single index. Yet some variables, such as teacher motivation, are inherently difficult to measure and thus are not used in any of these studies, so the large number of school variables used does not necessarily avoid bias due to omitted school and teacher characteristics. Moreover, in all four studies most school and teacher variables were not significantly different from zero, which reflects both low sample

sizes (163 students in Ghana, 355 in Jamaica, and about 250 in Brazil for the authors' preferred value-added regressions)[12] and high correlation among many of these variables.

A third problem is sample selection. In many developing countries some children never attend school, grade repetition is quite common, and a substantial fraction of children drop out of school after only a few years. Estimation problems can also arise due to the choices parents make regarding the schools their children attend, and actions parents may take to change those schools. Each of these studies attempted to address at least some of these problems. The Brazil study is the least satisfactory because of the assumptions used to achieve identification of the sample selection terms. It is not clear why the variables in the selection equation for on time promotion that are omitted from the achievement regressions (such as mother's education, number of students in the school, and the type of school) can be excluded from the latter regressions. The authors concede that their selection correction procedure "does rely heavily on the assumption that the probit errors are normally distributed" (footnote 103). The India study has similar problems. It appeals to the Brazil and Ghana studies for evidence that selection of students (in terms of "survival" to higher grades) does not matter. It does address selection into public and private schools but does not explain how the selection term is identified. The efforts to deal with selection bias are better in the Ghana and Jamaica studies. Both clearly explain the identification strategy (the identifying variables are characteristics of the school not chosen), and the Ghana study accounts for sample selection affects due to delayed enrollment and dropping out (using a similar identification strategy). In both cases controlling for sample selection has little impact on the results, which is consistent with the Brazil study (but not the India study). While this "regularity" may be good news, because it implies that bias due to school selection is probably small, results from more countries are needed before drawing general conclusions.

The fact that years of schooling, or grade in school, could be endogenous is a fourth problem. The India study appears to avoids it because all students in the sample are in the same grade, but there is still a sample selection issue regarding which children reach that

[12] Although the sample size in the India study was larger, with 902 students, they are concentrated in 30 schools, which limits variation in school characteristics.

grade. The Brazil study mentions it but does nothing. The Ghana study treats it as a sample selection problem caused by delayed enrollment; low grade repetition in Ghana implies that nothing further need be done. In Jamaica, delayed enrollment is not common, and grade repetition is moderate (a typical child repeats once during six years in primary school), so that study ignores this issue.

A fifth potential problem is measurement error in the regressors. None of the four studies addresses this, and in fact none ever mentions it. A plausible case can be made that most such errors are random, which implies underestimation of true effects. This may explain why in each study most of the teacher and school variables were insignificant. While it is not clear how serious a problem this is, future studies must address it, although how to do so will depend on the details of those studies.

A final issue is the specification of the dependent variable. All four studies used test scores in level form. A notable alternative, used only in the Brazil study, is the "value-added approach", which is motivated in part by fixed effects estimation that has long been used in analysis of panel data. The basic idea is quite simple. Suppose, for example, that one has test scores for a sample of children for two consecutive years, say, grades five and six. Assume as well that one has current data on the schools those children attend, but no data on the schools attended by those children when they were in grades 1-4. In addition, one has no data on the innate ability of those children nor on a host of other unobserved characteristics of children and schools. Such data can be used estimate value-added specifications, of which there are two variants.

The first approach uses the change in the test score over the two points in time as the dependent variable, with current child, household and school characteristics as the explanatory variables. The second uses the more recent test score as the dependent variable and include the prior test score as an additional regressor. The prior test score is almost certainly measured with error, so the second variant requires one or more variables that can serve as instruments. Under certain assumptions, the value added approach can reduce bias in estimates of the impact of school characteristics on student achievement. In particular, if the first test measures the impact of all child, household

and school variables that precede it in time, there will be no omitted variable bias due to lack of data (child, household or school) that pre-date the first test. In addition, if innate ability (or child motivation) is a fixed effect in a level regression, differencing test scores at two different period of time should difference out this, and any other, fixed effect.

However, the usefulness of the value-added approach is open to challenge. If one is examining student performance in primary schools, and school characteristics change slowly over time, the first advantage is minimal. Moreover, the information contained in, say, a fifth grade test score may have a higher signal to noise ratio than the information in the difference of the fifth and fourth grade scores. Only a comparison of a level specification with a valued-added specification will clarify this. More importantly, innate ability may not be a fixed effect. A more plausible specification is to interact innate ability with school quality, in which case it cannot be differenced out. Finally, all the other problems raised in subsection I.B still apply to the value-added approach. Thus, while value-added specifications are worth exploring (if the requisite data exist), findings based on them must be treated with caution.

This review of conventional studies leads to several conclusions. First, many studies suffer from multiple estimation problems and show only limited awareness of them. Second, recent studies have made some progress, but many problems remain. In particular, they use more sophisticated econometric methods, or at least show a clear awareness of many of the potential estimation problems, but they have not overcome all of these problems. Third, in my opinion there are two related problems that are difficult to resolve in conventional studies that attempt to estimate the impact of school characteristics on student achievement: omitted school characteristics and unobserved characteristics of children and their households. Regarding the first problem, although the Brazil, Ghana, Jamaica and India studies included large numbers of school characteristic variables in their regressions, there may be very important but hard to observe characteristics, such as teacher motivation, that are highly correlated with the variables that are observed, which will lead to biased estimates. Some results seem rather counterintuitive: for example, the most important single school characteristic in the Ghana study was leaking roofs. Perhaps the underlying relationship is that more motivated teachers, principals and parents were

more likely to keep the building in good repair. The inability to observe certain child and household characteristics, such as the child's innate ability and parental tastes for education, also leaves lingering doubts.

On a more positive note, if a large number of good conventional studies show that a specific school characteristic increases learning, there is a good chance that these studies are detecting a strong causal relationship, and policies could be based on such findings (the alternative being choosing policies without any evidence whatsoever). Yet there are only a small number of rigorous conventional studies. Fortunately, in the past few years several new approaches have been used to understand how school characteristics on student achievement. These are discussed in the following subsection.

D. New Approaches to Estimating the Impact of Policies on Education Outcomes

In recent years, both education researchers and economists have tried new methods to avoid the problems raised in subsection I.B. These can be divided into two types. The first retains the goal of estimating an education production function, or at least a reduced form version of it. The second abandons altogether attempts to identify specific school characteristics that make some schools better than others; instead it asks whether certain policies – such as vouchers, decentralized administration of public schools, or promotion of private schools – can raise students' cognitive skills.

Education production functions such as equation (2) in subsection I.B contain most of the information that a Ministry of Education wants to know.[13] These functions are technological relationships that show how much students learn when placed in certain types of schools with certain types of teachers (conditional on student and household characteristics). Education planners can use this information to assess the impact of each school and teacher characteristic on learning. Combined with cost data on these characteristics, they can "design" schools to maximize learning per dollar spent.

[13] One type of information not provided by education production functions is behavioral responses of households to education policies. As explained below, such information can be very useful.

Suppose that it were impossible to estimate an education production function using conventional econometric methods, due to the problems raised above. An equally useful, though probably more expensive, approach is to conduct a series of randomized trials, one per school characteristic, to evaluate the impact of changes in school and teacher characteristics on learning. Randomized trials are very common in medicine but very rare in the field of education.[14] Labor economists have conducted randomized trials to investigate the impact of welfare reform, guaranteed minimum incomes, and job training on labor force outcomes (see James Heckman, et al., 1999; Charles Manski and Irwin Garfinkel, 1992; and the special issue of the *Journal of Labor Economics*, 1993).

Results from several randomized trials on different school or teacher characteristics cannot be assembled into an education production function, because such trials provide only reduced form estimates of the impacts of those characteristics. Yet this is not a problem for policymakers; indeed, a limitation of knowing only the "true" education production function is that it does not incorporate households' behavioral responses. For example, suppose school quality increases in some way. One possible response of parents to higher quality is to reduce the time they spend helping their children with schoolwork. Such a behavioral response is not measured in an education production function, but would be measured in a randomized trial of that quality improvement (assuming that the randomized trial encounters no serious problems, an issue discussed further below).

The next paragraphs review two methods that, in principle, provide reduced form estimates of education production functions: randomized trials and natural experiments.[15]

---

[14] Three possible reasons why education researchers rarely use randomized trials are: a) most education policies are implemented at the classroom or school level, which greatly raises the costs of randomized trials – in contrast, most medical trials are randomized at the individual level; b) medical researchers have more experience with randomized trials because they often implement them using animals, since animal studies are much more relevant for understanding human health than for understanding education issues; and c) findings on human health in one country usually apply to humans generally due to common physiology, but results in education are typically much more specific to the local culture and school system. There may be other reasons, but I will not pursue them.

[15] A third approach is matched comparisons, which have been used to analyze U.S. job training programs (Heckman et al., 1997, 1998). Yet these methods offer only a modest

1. Randomized Trials. The basic idea of randomized evaluations of any kind is to compare two groups of observations that have no systematic differences other than one group received the "treatment" and the other did not. The simplest method is to sample a population of interest and randomly divide the sample into "treatment" and "control" groups. If this can be done without further complications – a big "if" – differences in the variables of interest across the two groups are unbiased estimates of the (reduced form) effect of the treatment.

In theory, randomized trials avoid all the problems discussed in subsection I.B. Random assignment of observations into treatment and control groups implies that both observed and unobserved characteristics of those observations are uncorrelated with treatment status. In econometric terms, the outcome of interest is the dependant variable and the only regressor is treatment status. That regressor is uncorrelated with everything in the error term because treatment status is uncorrelated with virtually everything. Another problem that randomized studies should resolve is measurement error; in any well-managed study treatment status should be measured without error.

In practice, randomized trials can have serious problems. First, child, household and school characteristics may change in response to the treatment. For example, if treatment schools are provided with abundant school supplies parental efforts to improve those schools (such as fund-raising activities) may decline. Even so, the only implication of this is that the impact of the treatment is a reduced form effect, rather than a structural parameter. As explained above, the former is often more useful for making policy choices. Yet structural estimates may also be of interest. Even if the reduced form effect on student learning is zero, a policy may still raise the welfare levels of parents and others. In the above example parents' welfare rises due to less time spent on fundraising.

---

extension of the conventional approach of controlling for observed school, teacher and child characteristics because they do not avoid the problem that observed and unobserved characteristics may be correlated. Moreover, I know of no studies on education in developing countries that use this method.

Another set of problems is sample selection issues, which are difficult to avoid. Parents of students in the control schools (or schools not included in the evaluation) may try to enroll their children in the treatment schools. This may affect the results by increasing class size (if class size affects the outcome of interest). This is not part of a reduced form effect because a nationwide adoption of the policy would not have this effect. In addition, children who transfer into the treatment schools may not be a random sample of the general student population. A related problem is that marginal students in the treatment schools are less likely to drop out (if the intervention raises student achievement), which will underestimate the impact of the policy on learning if comparisons are made based on all students currently enrolled in school. As discussed below, there are ways to reduce these problems, but they may not always work.

The first randomized trials of education policies in developing countries were done in the early 1980s by Stephen Heyneman, Dean Jamison and their collaborators. Jamison et al. (1981) conducted a randomized trial in Nicaragua in which 48 first-grade classrooms received radio mathematics instruction, 20 received mathematics workbooks, and 20 served as controls. After one year, students in the classrooms that received radio instruction scored more than one standard deviation higher on mathematics tests than students in the control group, and students in the classrooms that received mathematics workbooks scored about a third of a standard deviation higher. Both differences were highly statistically significant.

In the second study, Heyneman et al (1983) studied the first two grades of 104 primary schools in the Philippines. The schools were divided into three groups: 26 received mathematics, science and Pilipino textbooks at a ratio of one for every pupil , 26 received the same textbooks at a ratio of one for every two pupils, and 52 served as controls. Because textbooks were distributed to all schools in the 1977-78 school year, the control schools were evaluated in terms of student test scores in the previous school year (1976-77). Students in the two groups that received textbooks performed similarly, even though one had twice as many textbooks as the other; their test scores were about 0.4 standard deviations higher than those in the control schools (averaged over two grades and three subjects). These differences were also highly statistically significant.

The randomized experiments in Nicaragua and the Philippines were well designed and executed. Yet a potentially serious problem of both studies is sample selection and attrition. It is possible that enrollment increased in the treatment schools, which may have affected indicators of student performance. The direction of bias depends on the characteristics of the students attracted to those schools. If they were relatively weak students who otherwise would not have been in school, the bias is downward, but if they were strong students from other schools the bias is upward. A similar result on downward bias holds for attrition: if the intervention caused relatively weak students to stay in school longer the estimated effect of the program is biased downward. Another potential problem in the Philippines study is that students' test scores in the control group were collected one year earlier than those of the students in the treatment groups. It is possible that other differences in those two years could lead to biased results.

Did the results of these two studies lead to changes in education policies? The strong impact of radio instruction in Nicaragua may explain the expansion of educational radio to other Latin American countries (and a few countries in Africa and Asia) in the 1980s (John Newman, Laura Rawlings and Gertler, 1994). Ironically, Nicaragua abandoned education radio after the Sandinista government came to power in 1979; that government favored huge literacy campaigns, and its disputes with the U.S. government ended the USAID funding that had financed its education radio program. It is less clear whether the textbook results in Nicaragua and the Philippines led to policy changes; since most education officials would view this result as unsurprising, it may have had little effect. Unfortunately, no more randomized studies in education were done until the mid 1990s. The following paragraphs review recent studies done in Turkey, the Philippines and Kenya.

A recent study by Turkish educational psychologists is the only randomized study in a developing country not initiated by economists. Cigdem Kagitcibasi, Diane Sunar and Sevda Bekman (forthcoming) examined the impact of a mother education program on pre-school aged children. They considered three pre-school settings: "educational centers," which attempted to teach children specific skills; "custodial centers," which had no specific educational objectives; and children cared for at home. Within each group, about half of the children were three years old and half were five years old. For each of these six

age pre-school categories, mothers were randomly assigned to receive (or not receive) intensive "mother training". Ten mothers assigned to be trained "declined to participate" and were placed in the "no training" category. While this attrition is small it may lead to overestimation of the impact of the program if these ten mothers had lower than average tastes for their children's education. After four years, 25 (9%) of the original 280 mothers had dropped out of the program, leaving 255 children in the sample, 64 in educational centers, 105 in custodial centers and 86 cared for at home. No differences were found in an IQ test administered at the start of the program between the 25 children who dropped out and the 255 that remained. After two years of mother training, a variety sociological, psychological and achievement tests were administered. There were significant differences between the treatment and control groups for some outcomes, but not for others. The study found no significant program impact in terms of mathematics and (Turkish) reading ability, although the point estimates were positive, but did find a statistically significant positive impact on IQ scores and on "general ability" (spatial, numeric and verbal reasoning). This is puzzling because students with higher "ability" should be better at learning academic subjects. The magnitude of the estimated impacts is unclear because the study does not report the standard deviations of the test scores.

The Turkish study, while innovative, is open to several criticisms. The potential for bias caused by the ten mothers who declined training could have been avoided by using an instrumental variables estimation procedure, where actual treatment is instrumented by the original random assignment. This would measure the impact of the program on mothers who were trained, the "effect of the treatment on the treated". Retaining the ten mothers and regressing the outcome(s) of interest on the original random assignment would measure the effect of being offered the treatment. Other problems are harder to solve. First, the small sample of 255 children may explain why most of the results, which were in the expected direction, were insignificant. Second, there is no information on the costs of the intervention, which hampers cost-effectiveness comparisons with other studies. The program description suggests very high costs. Third, the mother training may have been implemented by highly motivated and highly trained individuals; implementing the program on a larger scale may draw less educated and less motivated trainers, reducing the program's effectiveness.

A second recent randomized study, Jee-Peng Tan, Julia Lane and Gerard Lassibille (1999), was also done in the Philippines. It examined four education policies: school feeding; multi-level learning materials (pedagogical materials for teachers); and combinations of each with "parent-teacher partnerships" (structured meetings between parents and school officials). Thirty schools were randomly assigned to five groups: five schools each for the four policy interventions and ten control schools. The authors examined dropout rates and student test scores after one year. They found almost no effects on dropping out: only the provision of multi-level materials had a significant impact (and only at the 10% level), reducing the dropout rate by about five percentage points. In contrast, most of the policies had significant impacts on test scores, though statistical significance varied with the estimation procedure used. Simple estimates that ignore selection bias due to differential dropout rates produced large impacts (as high as 0.87 standard deviations), although most were statistically insignificant. Correction for selection bias yielded significant effects more often, but with little effect on the point estimates. School feeding combined with parent-teacher partnerships most often produced sizeable and statistically significant impacts, ranging from 0.28 to 0.44 standard deviations for math, Filipino and English test scores. Multi-level materials with parent-teacher partnerships also had significant impacts, from 0.23 to 1.05 standard deviations for Filipino and English (but not math). School feeding alone had statistically significant impacts on English (and for math in one of three specifications), while multi-level materials alone had small impacts that were rarely statistically significant.

The authors conclude that combining multi-level learning with parent-teacher partnerships seems to be the most cost-effective policy, partly because of their regression results and partly because school feeding programs are expensive. Yet they recognize the imprecise and tentative nature of their results. The methods used to control for sample selectivity raise some doubts: for example, one of the identifying variables in the selection correction term is distance to the nearest school, but this could directly affect learning by causing children to be absent or tardy more often. Overall, the imprecision of these results and their sensitivity to estimation methods suggests that they be interpreted with caution.

The most recent set of randomized studies on education are those being conducted in Kenya by Michael Kremer, this author, and other collaborators. Six randomized trials have been conducted in rural Kenyan primary schools: a standard package of inputs (textbooks, school uniforms, and construction materials); textbooks only; block grants; flip charts; a package of teacher incentives; and treatment of intestinal parasites.[16] Results are currently available for four of these studies.

The first study in Kenya, Kremer et al (1997), examined the standard assistance package of a Dutch non-governmental organization (NGO). Fourteen schools participated, of which half were randomly chosen to receive assistance. There were no statistically significant impacts of the package on student test scores (English, mathematics, science, Kiswahili, geography/history/civics and art/craft/music), and the point estimates were small (less than 0.1 standard deviations). On the other hand, the program did reduce dropout rates. This study faced two serious problems. First, the sample size was small (in terms of the number of schools), which led to imprecise estimates. Second, the program increased enrollment in the treatment schools by an average of 35%, while in comparison schools enrollment declined by 10%. If higher class size lowers student achievement (and one study, discussed below, supports this hypothesis), the estimated impact of the program is biased downwards. The authors attempt to correct for this problem, but they have difficulty isolating the impact of the package from the impact of higher class size.

The second Kenya study, Glewwe, Kremer and Sylvie Moulin (2001), examines provision of textbooks. Rural primary schools in Kenya rarely provide textbooks; parents are expected to buy them, but few do. In 1995, 100 rural primary schools were randomly divided into four groups of 25 schools. In 1996, textbooks were provided to children in grades 3-8 in the first group of 25 schools. After four years, there is very little evidence of a sizeable impact of textbooks on the average test scores of students. Point estimates are usually 0.1 standard deviations or less, and in almost all cases impacts of 0.3 or higher can be ruled out. However, the authors do find evidence that textbooks benefited the better students. These overall results are at odds with the first randomized studies in Nicaragua

[16] A standard package of pre-school assistance is also being evaluated. Preliminary evidence indicates no effect of that package, but the final results are not yet available.

and the Philippines. Two possible reasons for the lower impact in Kenya are: 1. The teachers were not trained in the use of textbooks (extensive training was provided in the Philippines, but only minimal training was given in Nicaragua); and 2. The textbooks were too difficult for the average student in rural Kenya. The authors show that the typical median child in grades 3-5 could not read the textbooks provided (the official textbooks recommended by the Ministry of Education), although this was not the case for grades 6-8. Unlike the first Kenya study, provision of textbooks did not increase enrollment in the 25 treatment schools.

The third Kenyan intervention, examined in Glewwe, et al. 2000, focused on flip charts: large poster-sized charts with instructional material that can be mounted on walls or placed on easels. This intervention covered 178 primary schools, half of which were randomly selected to receive flip charts covering science, math, geography and health. Despite a large sample size and two years of follow-up data, the estimated impact of flip charts on students' test scores is essentially zero and completely insignificant. In contrast, several conventional OLS estimates, which may suffer from many of the problems described in subsection I.B, show impacts as large as 0.2 standard deviations, 5-10 times larger than the estimates based on randomized trials.

The most recent intervention in Kenya examines student health. Intestinal parasites (roundworm, whipworm, hookworm, and schistosomiasis) are endemic in rural areas of Kenya and many other developing countries. Medical research shows that high "loads" of intestinal worms lower scores on IQ tests, but almost no research has been done on their long-term impact on academic tests. Fortunately, treatment with albendazole every six months eliminates roundworm, whipworm and hookworm, and annual doses of praziquantel cure schistosomiasis. A sample of 75 schools was divided into three groups of 25 schools. The first group was treated in 1998, the second in 1999, and the third is a control group (to be treated in 2001). Analysis of two years of data by Edward Miguel and Kremer (2000) indicates that provision of albendazole and praziquantel increased student participation (fewer absences and reduced drop-out rates) but had no significant effect on test scores. In fact, the program slightly reduced test scores (by -0.04 standard deviations

after one year and –0.07 after two years; these are averages over English, math and science), but these impacts were statistically insignificant.

This more recent experience with randomized studies of education in developing countries provides several useful lessons. First, sample sizes should be quite large, at least 50 to 100 schools, to avoid imprecise estimates. Second, problems of differential selection into the initial sample (first Kenya study) and attrition (Philippines) across the two types of school are real possibilities; sound estimation methods that address these problems must be planned before data collection because they may require additional baseline data. Third, school outcomes should be followed for more than one year to see whether program impacts increase or fade over time. Fourth, a large amount of school data should be collected to check for other possible biases. An example of this is in the paper on textbooks in Kenya; it examined whether biases could be caused by reduced school fundraising, reduction in the purchase of textbooks by parents, and a greater tendency to promote students to the next grade, and found that none of these potential problems appears to overturn the result that textbooks had little or no impact.

2. Natural Experiments. Although well executed randomized studies can avoid many econometric problems, they can be very expensive to implement. An appealing (though rare) alternative is to find "natural" variation in a school characteristic that is uncorrelated with virtually anything else that determines child learning. Two recently published studies demonstrate what can and cannot be learned from such "natural experiments".[17] The first, by Case and Deaton (1999), examined educational outcomes in South Africa. The data used were collected in 1993, when government funding for schools was highly centralized, and blacks (people of African descent) had virtually no political representation of any kind. The authors argue that blacks did not control the funds provided to their children's schools, and that tight migration controls limited their ability to migrate to areas with better schools. They show that pupil-teacher ratios varied widely across black schools, and argue that this

[17] See Rosenzweig and Wolpin (2000) for a thorough discussion of "natural" natural experiments, i.e. natural experiments whose parameters of interest are identified by date of birth, twin births, gender of newborn child or siblings, and weather. The issues raised in that paper also apply to "less natural" experiments, and many are discussed below.

variation, combined with migration barriers and black South Africans' lack of control over their schools, generates a kind of natural experiment.

The South Africa study examines whether increased school resources lead to better educational outcomes. Most economists and other observers would probably expect the answer to be "yes", but Case and Deaton argue that some economists have claimed otherwise. They present several regressions that show the impact of school resources (primarily measured by student-teacher ratios) on years of completed schooling, enrollment and test scores. They find evidence that greater school resources increase all three outcomes. Specific findings are that decreasing the student teacher ratio from 40 to 20 (the approximate means in black and white schools, respectively) increases grade attainment by 1.5 to 2.5 years and raises students' reading test scores (conditional on years of school attendance) by the same amount as does two additional years of schooling (in contrast, there was no significant impact on math scores).

While the South Africa study has some data problems (e.g. the children tested were not a random sample of household members, and data from the Ministry of Education are not highly correlated – an R^2 coefficient of 0.15 – with the authors' community data), most readers would agree that, in principle, resources matter. The authors' evidence that some economists disagree is a statement by Hanushek (1995) that "providing more inputs...is frequently ineffective", which certainly allows room for inputs to be effective in some cases.[18] In addition to the blandness of the main result, two other criticisms can be made. First, even if blacks could not influence class size in their children's schools, certainly *someone*, presumably some government officials, made decisions that influenced class sizes in South Africa's black schools. If these decisions were influenced by education outcomes in those schools, they could yield biased estimates of the impact of class size (and, more generally, school resources) on those outcomes. This well known problem of endogenous program placement (see, *inter alia*, Mark Rosenzweig and Kenneth Wolpin, 1986). Second, the data cannot tell us how educational resources should be used; they

[18] Hanushek is more pessimistic on the impact of increased inputs in the U.S. and other developed countries; see, for example, Hanushek (1996).

provide no information to a Ministers of Education in developing countries on how to use any additional resources they may receive.

The other recent study based on a natural experiment is that of Joshua Angrist and Victor Lavy (1999a), who examine the impact of class size in Israel.[19] The natural experiment is a strictly enforced rule that limits class sizes to 40 or fewer students (a rule proposed by Moses Maimonides, a 12th century Talmudic scholar). The limits on class size determined by this rule vary in a highly nonlinear way with total enrollment in a given grade, providing an unusually credible instrumental variable to get around the problem that class size may be correlated with unobserved determinates of student learning. The authors use data from the early 1990s on a national test for Israeli 3rd, 4th and 5th graders. Most of the data are at the classroom level, so the analysis is at that level. The data are limited to Jewish public school students; private schools (mostly Jewish religious schools) are excluded due to their different curriculum, and Arab public schools (Arabs and Jews attend separate public schools) are excluded due to lack of data on "percentage disadvantaged" in Arab schools.[20] For each grade the sample is approximately 2000 classrooms from about 1000 schools.

The only explanatory variables used by Angrist and Lavy are class size, the percent of disadvantaged students in the school (averaged over all grades) and total enrollment for the grade. In most contexts this paucity of school variables would lead to omitted variable bias. Yet all one needs to obtain consistent estimates is an instrumental variable that predicts class size and is uncorrelated with the error term in the test score regression. The application of Maimonides rule is promising because it generates an oddly shaped relationship between class size and total school enrollment. In grades with an enrollment

---

[19] Another paper by the authors (Angrist and Lavy, 1999b) examines computer-assisted instruction in Israeli schools. The identification strategy in that paper is less appealing because there are no large discontinuity points in the function generating the instrumental variable. Moreover, most of the results based on that strategy are statistically insignificant.

[20] In fact, Arab schools could have been analyzed because the percent disadvantaged variable is not needed; it should be uncorrelated with the instrument constructed using Maimonides rule. The only reason to include that variable is to try to increase the precision of the estimates.

of 40 or less, class size will equal total enrollment. When total enrollment hits 41 the class must be split into two, so that class size is half of total enrollment for grades with 41 to 80 students. When total enrollment hits 81 a third teacher must be hired, so that class size is one third of enrollment for grades with total enrollment from 81 to 120. This "zig-zag" relationship between total enrollment and class size generated by Maimonides rule allows the authors to create an instrument for class size that is not highly correlated with total enrollment, so they can include total enrollment and its square as additional regressors.

Before examining the results, two comments are in order. First, as in randomized trials, the estimated impact of class size is not a production function parameter but a reduced form effect. When class size shifts abruptly due to application of Maimonides rule other classroom characteristics may also change, such as teaching methods or time spent on various activities. Yet from a policy perspective this information is very useful, as explained above. Second, even this estimation strategy may have problems. Some parents may know how Maimonides rule is applied, and those with high tastes for child education may transfer their children out of schools in which that rule leads to high class sizes. This can cause correlation between unobserved parental tastes for child education and the instrumental variable used to predict class size. The authors claim that this bias should be negligible (for example, Israeli parents would have to move to transfer their child into another school, or at least switch the child from a secular to a religious school), but there is no rigorous way to test for this problem.

Angrist and Lavy find a significantly negative impact of class size on the reading and mathematics scores of fifth graders. The estimated effects of a one standard deviation decrease in class size (reduction of 6.5 pupils) are increases in reading scores of 0.2 to 0.5 standard deviations and in math scores of 0.1 to 0.3 standard deviations (the range reflects differences in the sample and in the other covariates). The effects on fourth graders are less precisely estimated; sometimes they are significantly negative for reading scores, but for math scores the effects are all insignificant. For third graders all estimated impacts are insignificant; the authors suggest that this may reflect difficulty in measuring a presumably cumulative effect at lower grades. They also point out that testing conditions for the third graders were different from those for fourth and fifth graders.

These two studies of natural experiments in education in developing countries demonstrate both promise and pitfalls. In the South Africa study little was learned that school officials could use. While the Israeli study is probably the best study of the impact of class size on student performance in a developing country, it also highlights how much is left to learn. First, Israel is in many ways closer to a developed country than to a developing country. Second, the finding that class size matters is already assumed to be true by most officials in Ministries of Education, so it is unlikely that policies will change in response to this research. Third, this method probably cannot be applied to other countries because Maimonides' rule is used only in Israel. On a more positive note, both studies highlight what can be learned from a natural experiment and raise the intriguing possibility that more natural experiments are waiting to be discovered in developing countries. A very recent example is Esther Duflo's (2001) study of Indonesia; it is not reviewed here because it does not examine cognitive skills.

3. Studies on Private Schools and Decentralization of Public Schools. The implicit assumption thus far is that governments will use the estimates obtained to improve public schools. Another strand of recent research goes beyond this policy framework and instead considers decentralized management of public schools and private provision of education. This approach is due in part to dissatisfaction with estimates of education production functions, but also reflects doubts that governments have the right incentives to administer effective policies.[21] Thus the policies of interest are private provision of education and very decentralized public provision, often called "community schools". Interest in private schools does not imply no role for the government in education; private school advocates typically harbor serious doubts about the public *provision* of education, or at least centralized public provision, yet they may still support public *finance* of education, such as publicly provided vouchers to fund private or non-traditional schools.

[21] In fact, estimates of production functions can shed light on incentive issues. See Lant Pritchett and Deon Filmer (1999) for a discussion of how to use such estimates to investigate whether educational input choices favor teachers' interests over students' interests.

Early attempts to estimate the impact of private schools on learning were wholly within the production function approach. Researchers simply added a dummy variable indicating enrollment in a private school, and interpreted the coefficient as measuring the relative efficiency of private schools. This specification is appropriate if the increased efficiency takes the form of a constant multiplied by the production function; taking the logarithm of both sides of the production function yields such a specification. Yet this approach is too simplistic, because the dummy variable would also measure systematic differences in unobserved characteristics across public and private schools. For example, if private schools use some highly effective teaching method, and the data contain no information on teaching methods, the private school dummy variable would be positive. This positive coefficient does not necessarily imply that private schools are more efficient; it may simply indicate that private schools use a different set of inputs than public schools.

In theory, it is not even necessary to estimate a production function. Instead, one can look at the test scores of children in both types of schools, compare the costs of each type, and calculate the relative efficiency of each in terms of test score points per dollar spent. The obvious problem with this approach is that it ignores differences in student characteristics across public and private schools; child characteristics are likely to vary across these types of schools, and some key child characteristics may be impossible to observe. The ideal experiment to estimate the relative efficiency of public and private schools would allow school characteristics to vary within both types of schools but would randomly assign children to schools to avoid systematic variation in child characteristics.[22] Variation in school characteristics would lead to variation in expenditures per pupil in both types of schools. Regressing children's test scores on school expenditures, separately for both types of schools, estimates the efficiency of government and private schools in producing "achievement" at different levels of spending. School expenditures per pupil is the only explanatory variable needed; all school (and teacher) variables, observed and

[22] This "ideal experiment" is conditioned on the proportion of schools that are private. If that proportion increases, public schools may react by making changes that increase their efficiency, which would raise overall school efficiency but somewhat ironically would reduce the relative efficiency of private schools. Major public policy changes must account for these "general equilibrium effects" but they are beyond the scope of this

unobserved, are summarized by this variable. To my knowledge, such random assignment of students to public and private schools has never been conducted in any country, although something approaching it has been done in Colombia, as discussed below.[21]

Without random assignment of students, selection bias is likely; private schools may produce better educated students per dollar spent only because their students are more talented, or receive more parental support, than public school students. To correct for selection bias, observed variation in child characteristics (and parent characteristics and other household variables) can be entered as additional regressors. In principle, bias due to systematic differences in unobserved child characteristics across public and private schools could be corrected by using standard Heckman methods, namely by generating a selection correction term from prior estimation of the choice between public and private schools.

A fundamental requirement of standard selection correction procedures is that the selection correction term be identified, which can be done by either arbitrary distributional assumptions or exclusion restrictions. The former is almost never defensible, so empirical studies must rely on the latter, which means that they require variables that determine the choice between public and private schools but do not influence achievement once a school is selected. In theory, there are some obvious candidates. The characteristics of the school or schools not chosen should influence school choice but not academic achievement in the school chosen. Another candidate is school prices: even the price of the school chosen should not influence achievement in that school. A third candidate is distance to the school choices, which is a kind of price. Yet even these exclusion restrictions are questionable. If distance to the school the child attends is correlated with absences or tardiness, it may belong in the achievement regression. The price of the school attended is endogenous, as explained above. Even if it is exogenous it is likely to be correlated with unobserved characteristics of school quality and thus will be correlated with the error term in that regression. Characteristics of schools not chosen (including price and distance) are less open to criticism, but even here selective migration (moving closer to a desired school) can

---

paper. I know of no empirical work from developing countries that considers these effects.

cause problems. A final point is that standard Heckman methods to correct selectivity bias assume specific functional forms for the error terms of all equations; future studies should use more general methods, such as those suggested by James Powell (1994).

In fact, almost all studies to date have used a somewhat different approach. The school variables used include not only expenditures per student but also many other school and teacher characteristics. These studies typically use an Oaxaca-type decomposition to divide differences in mean test scores between public and private schools into differences in the means of observed characteristics and differences in the parameter estimates across the two kinds of schools. Assuming no estimation problems, the latter difference indicates the relative efficiency of private schools. Yet there is a fundamental problem with this approach; the parameter estimates for the additional regressors are likely to be biased due to omitted school characteristics, and biased parameter estimates will alter the two components of the Oaxaca decomposition. To see this, suppose that public and private schools vary in terms of an unobserved input. If this input is positively correlated with an observed input for which the mean is higher in private schools, it will increase the share of the first component of the Oaxaca decomposition (assuming that overestimation of the parameter on the observed variable is the same in both types of schools). On the other hand, if the mean is higher in public schools, it will increase the share of the second component. Note that this problem does not arise if the only school variable is expenditures per pupil, because that variable accounts for all variation in school and teacher characteristics that require expenditures.

A final problem in the literature comparing public and private schools is that measurement error in regressors is typically ignored. This applies to both school characteristics and child variables. The consequent biases in parameter estimates could also lead to biases in the Oaxaca decompositions, and could also bias estimates when the only school variable is expenditures per pupil. Standard instrumental variable methods could be applied; this may be feasible of the only variable is school expenditures but it

---

[23] In the U.S., the voucher program implemented in Milwaukee is also close to a randomized experiment; see Cecilia Rouse (1998) for a recent analysis of this program.

would be very hard to find instruments for a large number of school characteristics, especially given concerns of omitted variable bias.

Two of the best recent studies on the relative effectiveness of public and private schools in developing countries are Donald Cox and Emmanuel Jimenez (1991) and Kingdon (1996b). Yet these studies also exemplify the problems just discussed. Cox and Jimenez examine secondary schools in Colombia and Tanzania. They estimate selection correction terms using only family background variables and (for Colombia only) an "ability proxy", whether the child repeated a primary grade. The implicit exclusion restrictions are doubtful. Certainly, a child's ability should be included in the test score regression, and the same applies to many of the family background variables (such as parents' education). This study probably also suffers from omitted variable bias because the only school-level variables used are teacher salaries and the student-teacher ratio. All other differences between public and private schools go into the constant terms and thus are counted as measuring the relative effectiveness of the two types of schools. Finally, measurement error issues are not addressed.

Kingdon examines Indian students who are in "class 8", the final year of primary school. She uses the estimation method of Cox and Jimenez, except she divides schools into three types (public, private aided and private unaided) and then uses a multinomial logit model to estimate of school choice. The exclusion restrictions used have little theoretical basis (the author points this out in footnote 13); insignificant variables are dropped from the test score regressions but retained in the school choice regression. For example, mother's education is left out of the test score regression, excluding the possibility that educated parents help their children with schoolwork.. On a more positive note, Kingdon collected data on the per pupil cost for each school in her sample of 928 students in 30 schools. If one assumes that the observable variables on students account for all systematic differences in students across the different types of schools, the data provide a rare opportunity to compare the cost-effectiveness of public and private schools. Kingdon does this by predicting the test scores across all three types of schools for an average student and calculating the ratio of the cost over the predicted test score for each type of school (differences in learning across the three types of schools are estimated using

constant terms for each type and interactions of those constants with child and household variables). Private unaided schools have a ratio that is only about half of that for public schools and private aided schools, suggesting that fully private schools are much more efficient than public schools. However, while the sample selection terms have only marginal statistical significance, their presence or absence in the test score equations strongly affects the parameters of other variables in those regressions, which casts doubt on the results given the dubious identifying assumptions for those terms. Still, these large differences suggest a need for further research on this topic.

Another set of studies examines the relative effectiveness of public and private schools by using data from countries that have implemented voucher schemes, which provide government funds that students can use for private schools. In Chile, vouchers are in effect given to students in both public and private schools. The program began in the early 1980s, and by 1990 about 41% all students in primary and secondary were in private schools (compared to 22% in 1981). About 8% of these students were in private schools that charge tuition, which could not receive vouchers, leaving 33% in private schools that did not charge fees, which could receive vouchers. Two very recent papers on the Chilean experience are Patrick McEwan and Martin Carnoy (2000) and Alejandra Mizala and Pilar Romaguera (2000). The data used in both papers, however, have serious limitations. First, there is very little information about student characteristics. Second, the data are at the school level, so that impacts cannot be measured in standard deviations of the student distribution of test scores. Third, and most seriously, public schools must also compete for vouchers, which gives them a large incentive to be more efficient. Thus the comparison in both papers is between public schools and private schools when both compete for vouchers. Consequently, the data cannot be used to compare "ordinary" public schools (who need not compete for students) with private schools, nor to examine what happens when an "ordinary" school system is transformed into one where both public and private schools compete for vouchers (this change occurred in Chile in the early 1980s, and no data on student skills are available until the early 1990s). Both papers find few significant differences in the performance of public schools and private schools that compete for vouchers, but this is consistent with two very different hypotheses: a) the switch to vouchers had little effect on student performance in either public or private schools; and b)

the voucher system raised student performance by a substantial amount in both types of schools.

Colombia's voucher program provides a more useful comparison, because public schools did not compete for vouchers. Angrist et al (2000) examine the relative effectiveness of public and private schools, using a natural experiment. In Colombia, vouchers to attend private secondary schools were offered to over 125,000 students from poor urban neighborhoods from 1992 to 1997. In most communities where the demand for vouchers exceeded the supply, voucher eligibility was determined by a lottery, hence the natural experiment. Data were collected from 1600 applicants for the vouchers, stratified so that half were lottery "winners" and half were lottery "losers" (lottery losers were oversampled). Lottery winners were more likely to be in private schools than lottery losers (69% vs. 54%), which provides an instrumental variable for private school attendance that should be uncorrelated with virtually all determinants of student performance.

Angrist and his coauthors found that lottery winners completed more grades of schooling, primarily due to reduced grade repetition. However, this statistically significant effect is quite small, about one tenth of a grade. A potential problem with this result (acknowledged by the authors) is that lottery winners lose their eligibility if they repeat a grade. This gives private schools an incentive not to make lottery winners repeat in order to retain them as paying students. Thus the impact on completed grades of schooling may be an artifact of the program's design. The paper also examines test scores for a subsample who were tested. Of the 473 students asked to participate in testing, only 283 were tested. This refusal rate is high, but those tested do not appear to be a select group; lottery winners were no more likely to be tested than lottery losers. Reduced form estimates of the impact of winning the voucher lottery showed an impact of between 0.13 and 0.20 standard deviations (of the test score variable) for mathematics and (Spanish) reading and writing. Yet the low sample size led to higher standard errors; only the impact on reading was statistically significant, and only at the 10% level. Instrumental variable estimates in which test scores depend on student characteristics and a dummy variable for attending a private school, with private school attendance instrumented using lottery winner status, focused on progression in school, not test score performance. If a similar regression had been

estimated using test scores as the dependent variable, a significantly positive impact would be difficult to interpret because it may simply reflect more inputs and higher spending per pupil in private schools.

Finally, consider the very recent literature on decentralized management of public schools. The techniques used, and the associated problems, are almost identical to those in the literature on the relative efficiency of private schools. I know of only two papers, both by World Bank economists. First, Jimenez and Yasuyuki Sawada (1999) examined EDUCO schools in El Salvador, which are run by parent committees that can purchase school equipment and hire and fire teachers. The EDUCO program was not implemented in any randomized way, so the authors use standard selection correction techniques to avoid selection bias. Unfortunately, the data are from a sample of schools, so there are no data on schools not chosen (which is useful to identify selection correction terms). Jimenez and Sawada estimate achievement regressions that combine data from EDUCO schools and other public schools. They find that EDUCO schools outperform regular schools in terms of reading skills (by as much as 1.3 standard deviations) and daily attendance (by three to four days in the past four weeks). They conclude that decentralized management works by increasing the accountability of EDUCO schools to the local community. However, it is premature to draw such inferences for policy given the method used to control for sample selection bias. In particular, the selection correction term is primarily identified from arbitrary functional form assumptions; the only variables in the selection equation excluded from the equations of interest are district dummy variables, and there is no theoretical justification for this exclusion restriction. The estimates may also suffer from bias due to unobserved school characteristics.

The second paper on decentralization, Elizabeth King and Berk Ozler (2000), studies "autonomous" schools in Nicaragua. These schools have "directive councils" (the school principal, teachers, parents and even students) that select textbooks, set schools fees and hire and fire the school principal, tasks normally controlled by central authorities. The sample contains 1515 students in 80 autonomous and 46 traditional primary schools and 1455 students in 73 autonomous and 43 traditional secondary schools. The authors faced several formidable obstacles. First, the data indicate that autonomy varied widely within

both autonomous and traditional schools. In response, the authors formed two indices, *de jure* autonomy and *de facto* autonomy, to use in their regressions. Second, schools did not become autonomous at random but through a process that reflected, in part, the wishes of school officials and parents. Thus autonomy status and the two indices of autonomy are endogenous. The authors use instrumental variables, but the implicit exclusion restrictions are doubtful. For example, school enrollment is excluded from the test score regressions even though it may be correlated with unobserved school quality. Third, Nicaraguan schools experienced high attrition: transfers to other schools, dropping out and repetition (only non-repeaters were tested). To avoid sample selection bias, the paper uses the standard Heckman approach, although again using questionable exclusion restrictions. The IV results show no effect of *de jure* autonomy on student achievement, while *de facto* autonomy has a positive effect on primary school math scores (but not on primary school reading scores nor on secondary math or reading), although it is significant only at the 10% level.[34] Given the various estimation problems, it is premature to place much weight on this sole significant effect. Overall, the results of both studies of autonomy are intriguing and intuitively plausible, but more and better research is needed before making policy recommendations.

F. Lessons Learned and Suggestions for Future Studies

This review of the literature on the impact of educational policies on learning in developing countries clearly shows that much remains to be learned. Most of the conventional studies done in the last twenty years have serious problems. Some recent conventional studies have done a better job of grappling with fundamental estimation problems such as omitted variable bias, sample selection problems, and measurement error. Even so, results from conventional estimates must be treated with caution and should be regarded only as suggestive.

[34] The paper also presents regressions that focus on particular components of the *de jure* index; in one case a component was statistically significant at the 5% level for primary school (but not secondary school) reading and math scores.

Progress on measuring the impact of education policies is also hampered by the complexity of the education process and the wide variety in schools, teachers and students across developing countries. Variation in "similar" policies across countries often prevents general statements from being made. For example, textbooks appear to have had moderate but statistically significant impacts on learning in Nicaragua and the Philippines, but not in Kenya. One explanation for their ineffectiveness in Kenya is that the textbooks used were too difficult, at least in rural areas. Another possibility is that textbooks are effective only when teachers are trained in their use. A similar point holds for policies that seem to be highly successful. Education radio appears to have had a very large impact on learning in Nicaragua, but much of this success undoubtedly reflects the specifics of the radio programs produced. Such a policy will succeed in other countries only if the characteristics that made the Nicaraguan program successful are adopted (and appropriately adapted) in those countries.

My overall assessment of the literature is that most of what has been learned has been methodological in nature. First, the econometric problems inherent in conventional estimates of educational production functions are so daunting that it would be unwise to place much confidence in their results. Second, much more confidence can be placed in *well executed* randomized studies and natural experiments. Yet even these studies have many potential problems, such as nonrandom selection and attrition, inadequate sample sizes, and incorrect implementation of the intervention. All of these problems must be addressed in a convincing manner before making policy decisions based on their results.

Future work that attempts to estimate production functions should eschew conventional estimation methods and instead focus on randomized studies or natural experiments. Of these two options, the most promising is (well executed) randomized studies, since opportunities for credible natural experiments are likely to arise only rarely. More randomized studies are currently underway in Honduras, Mexico and Nicaragua, under the auspices of the International Food Policy Research Institute. Much can also be learned from studies that set aside the production function approach and address issues of school management, including differences between public and private schools. While most existing studies have serious weaknesses, the questions they ask are important and

should be pursued. If data are available on school costs, and on local schools that were not chosen, conventional methods can be used to assess the relative cost-effectiveness of public and private schools. Moreover, evaluation of decentralization and privatization policies can be done using randomized trials (as in Milwaukee) and natural experiments (as in Colombia).

II. Cognitive Skills and Labor Productivity

In both developed and developing countries, the cognitive skills children acquire in school play an decisive role in determining their standard of living as adults. The impact of cognitive skills on income is the most salient example: in almost every country, better-educated individuals have higher incomes. The most direct interpretation of this correlation is that the cognitive skills acquired in school are an important component of individual's human capital, and the return to that capital in the labor market leads to higher income.

Most economists would agree with this interpretation, and there is ample evidence to support it (such as Richard Murnane, John Willett and Frank Levy, 1995). Indeed, in a recent comprehensive examination of the causal relationship between education and earnings in developed countries, Card (1999) takes this interpretation for granted and never explores alternative hypotheses about the nature of human capital.[25] Economists often examine the relationship between years of schooling and income, yet more can be learned from examining the direct relationship between income and cognitive skills. First, positive correlation of cognitive skills with earnings, after conditioning on years of schooling, degrees obtained and measures of innate ability, casts doubt on other interpretations of the correlation between income and education, such as claims that such correlation reflects only sheepskin effects, individuals' innate ability, or learned acquiescent behavior (Samuel Bowles and Herbert Gintis, 1976). (Sheepskin effects are increases in income solely due to possession of a diploma or other certificate, as distinct from any effect of skills acquired from the education that the diploma or certificate represents.) Second, as discussed in

[25] Card does discuss whether degrees and certificates are rewarded in addition to (or instead of) rewarding human capital itself (the "sheepskin" hypothesis), but this is

Section I, in many developing countries schools are very ineffective in imparting cognitive skills to their students. Such countries would have weaker correlation between years of education and income. Data on cognitive skills and income can clarify whether this relationship is weak because the return to cognitive skills is low or because the impact of years of schooling on cognitive skills is low. Third, there are many different kinds of cognitive skills, and it is likely that some have larger effects on incomes than others. If the skills with the largest impacts could be identified, it may be schools should focus on those skills and de-emphasize others. Fourth, estimates of the relationship between cognitive skills can be used to estimate rates of return to investments in particular improvements to school "quality".

Despite these potential benefits, there has been little research in both developed and developing countries on the relationship between cognitive skills and income. The main obstacle is the paucity of data sets that include both the incomes and the cognitive skills of adults, although the situation has improved in recent years. This section examines the evidence from developing countries on the impact of cognitive skills on incomes, both wage income and income from self-employment activities.

A. Estimation Issues

In developed countries, research on the impact of education on income has focused on wage earners, who greatly outnumber the self-employed. In contrast, in developing countries the self-employed often outnumber wage earners; for example, the percentage of male workers who are self-employed or family workers is 62% in Bangladesh, 66% in Indonesia and 45% in Mexico (World Bank, 1998). The developed country literature focuses on two estimation problems, ability bias and measurement error in years of schooling (see Zvi Griliches, 1977, and Card, 1999). Lack of data on "ability" leads to overestimation of the impact of schooling on income if more able individuals go to school longer and ability has a direct positive impact on incomes beyond its indirect effect through higher years of schooling. In contrast, random measurement error in years

---

different from a hypothesis that claims that human capital is something other than productive skills acquired in school or from work experience.

of schooling underestimates the impact of schooling on income. Both estimation problems also arise if one regresses income on measures of cognitive skills (such as test scores) instead of years of schooling. Yet that the problem of ability bias may be reduced when cognitive skills are used because much of the impact of ability (conditional on years of schooling) probably works through the acquisition of cognitive skills. On the other hand, measurement error is almost certainly greater in cognitive skill variables than in years of schooling; tests of those skills certainly contain a substantial amount of noise due to poor test design, variation in "test taking ability", variation in testing conditions, and random fluctuation in individuals' health or attentiveness on the day of the test.

Additional estimation problems arise in developing countries because most workers are self-employed. First, the division of the labor force into wage earners and the self-employed is certainly not random, so that regressions that include only wage earners may suffer from sample selection bias. Second, it is difficult to calculate the incomes of self-employed workers whose economic activity involves many people, such as several family members working together. One can collect data only on the income of the "team", but analysis at the team level is problematic because education (whether measured by years of schooling or by test scores) often varies among the team members, raising the issue of how to measure the team's education. In practice, two or three approaches should be tried to check whether the results are sensitive to the method used (see Dennis Yang, 1997). Third, measuring hours of work of the self-employed is difficult because their hours vary day by day and from season to season. Fourth, in countries where wage earners are a small percentage of the labor force a large proportion of them may work for the government. Since governments face few economic forces that dictate that employees be paid their marginal products, the returns to education among government workers primarily reflect government pay scale policies and may be only weakly correlated with worker productivity. In such cases separate estimates should be done for both government and private sector workers, and analyses should focus on the latter.

While these additional difficulties hamper analysis of data from developing countries, there is one useful advantage. In general, sheepskin effects do not arise for self-employed workers (Wolpin, 1977, first made this point). More generally, the

incomes of self-employed workers are closely tied to their actual productivity. Thus data on self-employed provide an alternative method to test whether returns to education are primarily sheepskin effects. More broadly, data on the self-employed allow one to estimate more direct relationships between education and worker productivity than do data on wage workers.[26]

B. Recent Empirical Work

1. Wages. Most research on the relationship between income and cognitive skills in developing countries has focused on wage earners. The first such study, Maurice Boissiere, et al (1985), examined urban wage earners in Kenya and Tanzania. They authors investigated whether the positive correlation of schooling with wages primarily reflects differences in workers' cognitive skills, the alternative hypothesis being that this correlation primarily reflects the impact of innate ability or "credentialism" (sheepskin effects). Their data had the standard variables used in wage regressions, plus scores on three tests, a reading test, a mathematics test and the Raven's test of abstract thinking ability (the same test used in several studies discussed in Section I). The authors regressed annual earnings on work experience (years since leaving school), years of schooling, the sum of the reading and literacy tests (their measure of cognitive skills), and the Raven's test. The impact of the cognitive skill variable was almost always significantly positive, while years of schooling and the Raven's test were almost always insignificant. When the test score variables are omitted, the coefficient on years of schooling is much higher and often statistically significant. The authors claim that these results demonstrate that education raises wages by providing workers with cognitive skills, and do not support the alternative hypothesis that education primarily reflects innate ability and/or sheepskin effects.

While some aspects of the Kenya-Tanzania study can be criticized, it is difficult to see how its shortcomings could overturn its main findings. For example, there is likely to

[26] A way to measure worker productivity even more directly is to examine income from piece-rate work; an example is Andrew Foster and Rosenzweig's (1993) study of the Philippines.

be random measurement error in the achievement test scores, but correcting for such error would tend to increase the associated coefficient. On the other hand, measurement error in the Raven's test may underestimate the true impact of "innate ability", and any measurement error would be magnified by the fact that this test is not really intended to be a measure of only genetically inherited intelligence. Another criticism is that the sample mixes public and private sector workers, yet it is hard to see how including public sector workers would overestimate the impact of cognitive skills on wages. Similarly, one would expect "sheepskin" effects to be lower once public sector workers are excluded. Finally, one could also fault the paper for ignoring sample selectivity, but here again it is hard to imagine how selection bias could drive the results. Overall, this paper provides credible evidence to support its main conclusions.

A more recent study of Ghanaian workers by Glewwe (1996) is similar to the Kenya-Tanzania study, but it distinguishes between government and private sector workers and attempts to control for sample selection. It also finds that cognitive skills, as opposed to years of schooling per se, are the fundamental determinants of wages and also uncovers no evidence that innate ability (measured by the Raven's test) directly determines wages. Of course, this study also has shortcomings. It did nothing to correct for measurement error in the explanatory variables, and one could quarrel with its approach to correct for sample selectivity (the identifying variables in the selection correction term – marital status, family size, and parents' occupation – may directly affect wages). Also, the small sample size led to imprecise results. Yet again there is no reason to think that these flaws determine the results. A particularly noteworthy aspect of this study is its use of results on learning in Ghana from Glewwe and Jacoby (1994) to estimate rates of return to specific school quality improvements: the rates of return to those interventions were often higher than those from an additional year of schooling. Specifically, the estimated rate of return was 6-7% for providing textbooks, 15-25% for providing blackboards, and 13-24% for repairing classrooms with leaking roofs, while the rate of return to an additional year of schooling *at the current level of school quality* was 4% to 6%.[27] While these estimates are imprecise, they are much more useful to

[27] See Glewwe (1999a) for a more detailed analysis of rates of return to school quality.

policymakers than are rates of return to an additional year of schooling, the typical "output" of wage regressions.

The focus shifts to rural areas of Pakistan in a paper by Alderman, et al. (1996b). The authors directly confront several estimation problems. They control for sample selection bias and use instrumental variables for cognitive skills, years of schooling, and years of work experience.[28] Another notable aspect is that the wage regressions include measures of health status as explanatory variables, although they are never statistically significant. Their results from Pakistan are strikingly similar to those in Kenya, Tanzania and Ghana. "Ability," again measured by the Raven's test, has no statistically significant impact on wages, and when both cognitive skills and years of schooling are used as regressors the former is often statistically significant while the latter never is.

Evidence from Morocco is found in Angrist and Lavy (1997), who focus on that nation's "Arabization" policies (Arabic replaced French as the language of instruction in middle and secondary schools). In Morocco, language skills explain wages even after controlling for years of education yet, unlike the three studies just discussed, years of schooling still has strong explanatory power. This does not necessarily imply that schooling rewards workers in ways other than increasing their skills; since the tests cover only some skills, other skills may be picked up by years of schooling. Also, many of the workers in the data may be government employees, whose wages can reflect factors other than their productivity. A final remark is that none of the variables in the Moroccan data can be interpreted as measuring innate ability, so the results shed no light on the relationship between ability and wages.

The most recent study of the relationship between cognitive skills and wages in a developing country is Peter Moll's (1998) study of South Africa. The data available limited the options for dealing with several estimation issues. For example, no instrumental variables could be found to correct for measurement error in cognitive achievement, and there were no measures of learning ability. Yet the results are similar

[28] The use of instrumental variables may be more important in terms of correcting for measurement error than in terms of removing simultaneous equations bias.

to those discussed above. In particular, cognitive skills are strongly associated with wages even after conditioning on years of schooling. In turn, the impact of years of schooling diminishes, although it retains statistical significance, when the cognitive skill variables are added. As in Morocco, this last result does not necessarily imply that something other than cognitive skills raises worker productivity. First, the tests used in South Africa were at about a third or fourth grade level and thus do not measure cognitive skills at the secondary and tertiary levels. Second, years of schooling may reflect other types of cognitive skills. Third, the South African data probably include many government workers, whose wages may not closely reflect their productivity.

These five studies of wages and cognitive skills in developing countries provide two general findings. First, and most important, simple measures of basic cognitive skills have explanatory power beyond that given by years of schooling, and in three of five cases adding those variables led to insignificant explanatory power for years of schooling (particularly in the studies that include few or no government workers, Ghana and rural Pakistan). Second, "innate ability" does not affect wages after conditioning on years of schooling and cognitive skills. Although the Raven's test may measure innate ability imperfectly, one can still interpret its lack of predictive power as indicating that ability has no sizeable direct impact on wages.

While some degree of confidence can be put into these two findings, more research is needed to confirm them in other settings and to investigate other hypotheses. The sample sizes in these studies were small and often indiscriminately mix government and private sector workers. Only two of five studies, those on Morocco and Pakistan, address problems of measurement error in years of schooling and in the cognitive skills variables. Finally, the tests used were rather narrow and in some cases overly simple. For example, the South Africa mathematics and reading tests were based on six questions each. A final criticism is that these studies focus on wage workers even though self-employment is much more common in all of these countries. The following paragraphs examine three studies that focus on the self-employed.

2. Self-Employment Income. As mentioned above, self-employment income is very closely tied to productivity. Regrettably, only two published studies have examined the impact of cognitive skills on self-employment income in developing countries. The first is by Dean Jolliffe (1998), who estimates the impact of mathematics and reading skills on the agricultural, non-agricultural, and total income of 1388 Ghanaian households. He finds that cognitive skills raise non-farm income and total income, but not farm income. This suggests low returns to numeracy and literacy in agricultural activities in Ghana, which induces households with relatively high skills to move out of farming and into non-farm activities. To the extent that farming in Ghana consists mostly of routine activities that are unaffected by recent technological advances, Jolliffe's findings are consistent with Rosenzweig's (1995) conjecture that education raises productivity by increasing individuals' access to, and their ability to process, new information.

Jolliffe's paper is quite innovative; perhaps the main criticism is what it did not examine. It never used the Raven's test data to see whether they have any explanatory power beyond that provided by cognitive skills and years of schooling. Also, it does not investigate whether schooling becomes insignificant when skills variables are added. Finally, it does not examine whether agricultural productivity per hour of work, rather than total income, was affected by cognitive skills, since the absence of an impact on total agricultural income may reflect decreased time spent in that activity. (Similarly, part of the positive impact of skills on non-farm income may reflect more hours in that activity.) On the other hand, Jolliffe rigorously addresses problems of selection bias due to the fact that only some households engage in agricultural activities.

The only other study of the impact of cognitive skills on self-employment income is Wim Vijverberg (1999), who examined the same data used by Glewwe and Jolliffe. The author examined 1074 household enterprises in Ghana. Unlike the studies on wages and on total household income, he finds only weak evidence that schooling, measured either by years of school attendance or by cognitive skills, affects income from such enterprises. In fact, the impact of cognitive skills is often weaker than that of years of education. Yet there is one point of agreement with the wage studies: innate ability as measured by the Raven's test has no significant impact on income from non-farm self-

employment activities. Vijverberg concludes that impact of education on non-farm self-employment income is complex and probably varies by the type of business.

Vijverberg's analysis has no serious flaws, but unlike Jolliffe he does little to account for sample selection bias. Another difficulty, which is carefully addressed, is that the data on household self-employment income are quite noisy. In particular, they are based on a four-week recall period, a brief space of time that cannot be lengthened because survey respondents' memories on self-employment incomes would become even less reliable. Indeed, the best way to measure incomes from non-agricultural activities is a complex problem, one that led to a separate paper (Vijverberg, 1992).

While the impact of cognitive skills on income from self-employment is clearly an important topic, there is very little research on it. The only two published studies use the same data from a single country. While data on the self-employed avoids problems of sheepskin effects, it introduces other problems, such as how to measure the education level of a group of people and very noisy income data. It is premature to draw any general conclusions from only two studies. Much more research is needed to understand how cognitive skills affect self-employment income.

C. Gaps in the literature and suggestions for future research

The five studies that used wage data yield two tentative conclusions. First, cognitive skills directly affect wages, and may be the most important determinant of worker productivity. Second, "ability" does not appear to affect directly the productivity of either wage workers or the self-employed after controlling for years of schooling and cognitive skills. Of course, these results raise as many questions as they answer. Future research should go beyond simple tests of mathematics and reading to examine skills such as scientific, agricultural and health knowledge, and abstract thinking skills. It is very likely that the effects of different skills will vary by occupation, and indeed may determine which occupations are chosen. Finally, more research on the self-employed is urgently needed, because they are the majority of workers in most developing countries and the evidence thus far is limited to a single country (Ghana).

Future studies must also address problems of measurement error in cognitive skills variables. The best approach depends on the type of measurement error. Random errors by respondents in answering individual questions on a given test are easy to deal with. For each test, the questions can be arbitrarily divided into two sets, one of which can serve as an instrumental variable for the other. Yet this will not work for measurement error in the test as a whole, such as bad testing conditions on the day of the test and random inattentiveness or illness of respondents on that day, since the same measurement error would be contained in both sets of questions. Ideally, people should be tested twice, on different days and under different conditions. Although this will increase costs, the benefits are quite high. This may also reduce respondents' cooperation, and thus lower participation rates, but in developing countries refusal rates are much lower than in developed countries. Another alternative is to "test the test" on a random sample of people to estimate directly the extent of measurement error in the test.

As with analysis of the impact of school and teacher characteristics, future research on the impact of skills on incomes will require data collection expressly for that purpose. Although planning and executing new household surveys is expensive, the cost is almost certainly very small compared to the potential benefits. On a more cost-conscious note, there does not appear to be a strong case for randomized trials or natural experiments to analyze the impact of cognitive skills on incomes. Economists have many years of experience estimating wage and income equations and have developed many methods to overcome, or at least minimize, biases that arise on a variety of fronts. Indeed, it would be very hard to conduct a randomized trial that generates random variation in cognitive skills in the general adult population. Perhaps a randomized trial of school inputs could do this for a population of young adults, but one would have to wait at least 5-10 years before data are available for analysis; to my knowledge this has yet to be done.

III. Cognitive Skills and "Non-Economic" Outcomes

Schooling also affects many socioeconomic outcomes other than income, such as health status, migration, marriage prospects, and fertility. Presumably, many of these

effects operate through cognitive skills acquired in school. Yet almost nothing is known about which skills have the strongest influences on these outcomes. Research on the impact of skills on "non-economic" outcomes broadens the scope of the skills considered, moving from mathematics and reading to other skills and types of knowledge. For example, fertility and health outcomes may depend more on an individual's knowledge of health and science than on his or her mathematical ability. Yet reading and mathematics skills may still matter: one may need to read the directions on a medicine bottle and then know enough arithmetic to measure out correctly the prescribed dose of medicine.

There is very little literature on the impact of cognitive skills on non-economic outcomes in developed countries, and even less for developing countries. However, in recent years a few studies that use data from developing countries have been published. These are reviewed in the next sub-section.

A. Recent Studies: Fertility and Child Health

Four recent studies have examined the impact of cognitive skills on non-economic outcomes in developing countries. Two considered the impact of women's education on their fertility, while two examined the impact of mothers' education on child health.

Estimating the causal impact of education on fertility is a difficult task, since both are endogenous variables. An example of "reverse causality" is that teenage girls in developing countries who become pregnant typically drop out of school. The two studies examined here exemplify this problem. Duncan Thomas (1999), using the same South African data set analyzed by Moll, finds a strong and statistically significant negative correlation between years of schooling and children ever born among South African women, even after controlling for several other variables. When test scores on mathematics and reading comprehension are added, the latter is statistically significant and the coefficient on years of schooling declines by one third, though it is still significant. This suggests that at least part of the correlation between schooling and fertility works though cognitive skills. Since the test scores measure skills at about the third or fourth grade level (as explained by Moll), tests covering a broader range of skills

would probably have reduced the impact of years of schooling even further. Thomas makes no claim to have found a causal relationship. He speculates that reading skills improve women's ability to gain access to and assimilate information, and presents evidence consistent with this hypothesis, but cannot go further with the data at his disposal.

The second study on fertility is by Raylynn Oliver (1999), who analyzes the same Ghanaian data used by Glewwe, Jolliffe and Vijverberg. She also finds a strong and statistically negative impact of years of schooling on fertility (in terms of children ever born), though she is more willing than Thomas to interpret this as a causal relationship. When test scores for reading and mathematics are entered, her findings are very similar to those of Thomas – only reading scores have significant negative effects, and when tests scores are added the years of schooling coefficient declines by about one third but remains statistically significant. Another interesting finding is that "ability," measured by the Raven's test, has no significant impact on fertility. While Oliver too quickly ascribes causal impacts to her findings, the similarity with Thomas' results is striking. The absence of an effect of the Raven's test is also noteworthy, since it suggests that innate ability has no effect on fertility apart from its indirect impact through increased cognitive skills.

Finally, consider the impact of mothers' education on child health. Many studies find a strong and significant impact of maternal years of schooling on child health (see Behrman, 1990). There are several possible mechanisms that could explain this relationship. Perhaps education directly increases mothers' knowledge of health and health care procedures. Alternatively, basic literacy and numeracy skills may be more important than health knowledge per se. A third possibility is that schooling reduces women's adherence to traditional cultural practices, making them more receptive to modern health care treatments. Finally, increased maternal schooling may improve children's health outcomes by increasing household income.

Distinguishing between these different pathways is difficult because there are almost no data sets with detailed information on all these potential effects of schooling.

Perhaps the only such data is the Morocco survey used by Glewwe (1999c) to investigate these issues (the same data used by Angrist and Lavy). For 2171 Moroccan households, adult household members were given tests on reading and writing (both Arabic and French), mathematics, "general knowledge", and health knowledge. The test of health knowledge contained five questions on topics relevant to child health: vaccinations, treating infections, polio, diarrhea and safe drinking water. After excluding households without young children and observations with missing data, a sample of 1495 children age 0 to 5 years remained. Child health was measured by height-for-age. Glewwe's analysis of these data led to two conclusions. First, health knowledge appears to be the most important skill that mothers need to care for their children. Second, Moroccan mothers do not directly acquire health knowledge in school; indeed, it is not part of the standard curriculum. Instead, they acquire it indirectly by using the literacy and numeracy skills acquired in school. These findings suggest that Moroccan schools should seriously consider adding basic health education to the primary and secondary school curriculum, since such a change could significantly improve child health.

While the findings of the Morocco study are potentially very important, they should be treated with caution. First, the study found evidence that health knowledge should be treated as an endogenous variable and, as always, one could quibble with the instrumental variables used. For example, the instruments for the mother's health knowledge include the presence of radios and televisions in the household, but one could imagine more direct connection of these variables with child health (e.g. small children in households with televisions have less contact with other children). Second, the health knowledge test contained only five questions, and thus gives little guidance on the content of a new health curriculum in primary and secondary schools. Third, the evidence is based on only one country. Similar studies using data from other countries are needed to check the robustness of the finding that mothers' health knowledge is the key pathway by which maternal education raises child health.

The other study of mother's cognitive skills and child health is by Glewwe and Jaikishan Desai (1999), who also used the now familiar data from Ghana. They examined two dependant variables for a random sample of 1107 Ghanaian children age 0

to 5 years: height-for-age and weight-for-height. Ability (the Raven's test) never has a statistically significant impact. In the height-for-age regressions, none of the mother's education variables – years of schooling, mathematics score and reading score – was statistically significant, an unexpected result. In the weight-for-height regressions mothers' mathematics scores are generally significant, although not very precisely estimated. Neither years of schooling nor the reading score are ever significant. This suggests a role for mathematics skills, but the Morocco study suggests that if data on health knowledge had been available the impact of mathematics skills would become insignificant.

The study by Glewwe and Desai has several weaknesses relative to the Morocco study. First, there are no data on health knowledge, which the Morocco study suggests is critical. Second, the Ghana study did not use instrumental variables for the test score variables, which (if credible instruments can be found) could have reduced problems of measurement error. Third, the insignificant impact of education on height-for-age is puzzling, since that measure of child health generally has a higher signal-to-noise ratio than does weight-for-height. On a more positive note, the insignificance of the Raven's test suggests (but hardly proves) that innate ability alone will not improve child health.

B. Gaps in the Literature and Suggestions for Future Research

The literature on the impact of cognitive skills on "non-economic" outcomes is very new and very small, leaving many gaps. Probably the only tentative conclusion to draw is that there is no evidence (yet) that ability, at least as measured by the Raven's test, directly affects fertility or child health. Yet this is based on only two studies that use the same data from the same country. The most intriguing finding is the impact of health knowledge on child health, but this result needs confirmation using data from other countries before drawing general policy conclusions.

Future research should go in several different directions. First, other outcomes are of interest, such as adult health, marriage outcomes and perhaps political participation. Second, there are undoubtedly other kinds of skills – and values – acquired from formal

schooling that affect non-economic outcomes. Third, problems of measurement error in the tests must be addressed, as explained in the discussion of income outcomes. In the only study of the four that did so, the Morocco study, the coefficient on health knowledge increased several fold when that variable was instrumented. This leads to the issue of endogeneity of test scores; in the Morocco study the findings suggest that a "sickly" child increases a mother's health knowledge, and the joint determination of fertility and education outcomes is a formidable obstacle in any study of fertility. Fourth, all four of these studies used data collected especially for research purposes, and future progress is unlikely unless special data collection efforts are made.

IV. Summary and Concluding Comments

Developing countries spend hundreds of billions of dollars each year on education, and there is ample evidence that these funds are spent inefficiently. More effective use of these funds could increase the rate of human capital accumulation, which would increase incomes and, more generally, raise living standards in these countries. This paper posed three questions concerning the determinants of cognitive skills, and the impact of those skills on income and other socioeconomic outcomes. This section summarizes the evidence on each question and makes suggestions for future research.

The first question was: What school policies are most cost-effective in producing students with particular cognitive skills, such as literacy and numeracy? Until recently almost all empirical studies that addressed this question estimated production functions for cognitive skills. That is, they regressed students' test scores on a variety of school, household and child characteristics. It is now clear that this approach has serious shortcomings. Biased parameter estimates can arise due to omitted variable bias, endogenous program placement, sample selection bias, and measurement error in the explanatory variables. Can these problems be overcome? Most studies that attempted to control for sample selection bias have found little or no evidence of substantial biases. The situation is less clear for bias due to endogenous program placement, since almost no studies of education in developing countries have examined this problem. Unfortunately, the problem of omitted variable bias is likely to be severe, which explains why different

studies have produced very different results. Even worse, it is very difficult to overcome this problem because schools differ in so many ways, many of which are difficult to observe under even the best of circumstances. Finally, it is likely that measurement error problems lead to substantial biases, and there is no simple solution to this problem. Thus, all estimates of production functions for cognitive skills using conventional econometric methods should be regarded as suggestive, not definitive. This problem has led to new approaches, and to new questions that go beyond the production function approach.

Another approach to studying the impact of school policies on child learning is to generate or find random variation in school characteristics and compare cognitive skills across schools with different levels of those characteristics. This can be done by conducting trials that randomly divide schools into those that participate in a new policy and those that do not. This was first done two decades ago in Nicaragua and the Philippines, but more recently it has been initiated in several other developing countries. Similarly, one can search for instances in which unintended random variation in school characteristics generates a "natural experiment" that randomly divides schools or students into groups that do and do not participate in a given education policy. These are even rarer than randomized studies, but they are based on the same principle. The small number of studies completed thus far limits how much as been learned. The explicit lessons so far are: 1. In two of three cases textbooks or workbooks increase students achievement by 0.3 standard deviations; 2. In the one case where textbooks did not provide significant results (Kenya), the problem may have been that the textbooks were too difficult or that teachers were not trained in their use; 3. Education radio may be a highly effective method to raise student achievement in mathematics and science; and 4. Reducing class size does increase child learning, at least in a country with a relatively high level of income (Israel).

These policy findings are admittedly scant compared to the need for advice faced by Ministries of Education in developing countries. Yet recent experience with randomized trials provides information of another sort that will ultimately prove more useful than a longer list of what "works" and "does not work" in various developing countries. This is because developing countries differ enormously in their level of development, their

culture, and their education systems. Also, most education policies are defined in terms of a long list of characteristics that are difficult to summarize in a single sentence, or even a single paragraph. Consequently, one will rarely find that a "policy" defined in a simple word or phrase will be "good" for most or all developing countries. Instead, each country must rigorously test different policies, and indeed different versions of the "same" policy, to see which version of each policy, if any, works well for that country. This is the real promise of randomized trials (and natural experiments, which will remain rare) for developing countries. The contribution of this literature is not a list of what works and what does not, but advice on how each country can determine what works best for it. Such advice is given in detail at the end of this section.

A final aspect of the first question departs from production functions and instead asks what kinds of overall management policies best raise students' test scores. There is little reliable evidence so far; the best study of the relative performance of public and private schools (the study of Colombia's voucher program) provides evidence in favor of vouchers that can be used to attend private schools. However, these results do not necessarily imply that private schools are more cost-effective. On a more methodological note, randomized trials and/or natural experiments may not be needed to study this issue. The reason is that the only characteristics one needs regarding schools are whether they are public or private, and how much they cost to operate. All other school characteristics are irrelevant, so there is no problem of large numbers of unobserved school characteristics. The other problem confounding conventional estimates, selection bias due to unobserved child and parent characteristics, can be addressed using standard selection correction methods if one has characteristics of other schools that could have been attended, which provide exclusion restrictions that identify the selection correction term(s). In principle, this approach applies for any general management intervention, such as decentralization or charter schools, that allows one to label schools as one type or the other.

The second question was: What is the relationship between schooling, particularly cognitive skills acquired in school, and labor productivity? Recent studies on wage workers and the self-employed have provide two consistent findings. First, cognitive

skills appear to be directly responsible for part, if not most, of the impact of schooling on labor income; even after controlling for years of schooling and measures of "innate" ability, cognitive skills almost always have statistically significant impacts on income. Second, there is no evidence that innate ability, at least as measured by the Raven's test, has a direct impact on labor income after controlling for cognitive skills and years of schooling, although it does have a strong indirect impact by enabling individuals to acquire cognitive skills.

There are also some important methodological lessons. First, attempts to use test scores as regressors require an instrumental variable to correct for measurement error in the test score variable. Second, in developing countries there can be serious sample selection bias because often less than half of income earners are wage workers. These are difficult problems; serious thought is particularly needed on what variables, if any, determine selection into an occupation type but do not determine income from that occupation. Third, there is less need for randomized trials to determine the impact of skills on labor productivity and income, which is fortunate because such trials would be difficult to conduct.

The third question was: What impact does schooling, especially cognitive skills, have on socioeconomic outcomes other than labor productivity? There is very little evidence on this issue. One tentative finding, based on only two studies from a single country, is that a mother's innate ability does not affect her children's health or her fertility after controlling for years of schooling and cognitive skills. Future studies are needed to assess the robustness of this finding. The other finding, based on only one study, is that mother's health knowledge, as opposed to other knowledge or skills, seems to be the key contribution of education to child health. Again, further evidence is needed before making general policy recommendations. Finally, two methodological lessons for the second question also apply to the third. First, when using test scores as explanatory variables one must use instrumental variables; the impact of instrumenting on the Moroccan health study demonstrates this. Second, randomized trials are not essential to study this question.

The standard concluding call for more research applies *a fortiori* to research on schools, skills and socioeconomic outcomes in developing countries. Thus I conclude with advice to future researchers, beginning with advice that applies to studies of all three questions. First, in almost all cases one must collect original data because existing data are unlikely to be adequate for the task. For studies of the impact of school characteristics on students' learning, this usually implies organizing a randomized trial, while for all three questions it typically implies, at minimum, collecting data on cognitive skills to complement existing data. This can be expensive, but the costs are trivial compared to the hundreds of billions of dollars spent on education each year. Moreover, in some cases the additional data collection was relatively inexpensive. For example, the marginal cost of supplementing the 1988-89 Ghana Living Standards Survey with data on cognitive skills was only $100,000. If education systems already collect test score data, even the cost of randomized trials is not very high; the study of flip charts in Kenya used existing test score data, and the additional cost of randomly providing the flip charts was only about $50,000. If new test score data have to be collected, and the students are followed for several years, the costs can be higher. The Kenya project that included the textbook, grant and teacher incentive interventions cost about $450,000, and the same figure applies to the Kenyan deworming intervention. Note finally that Ghana and Kenya have low labor costs; total costs in middle income countries may be substantially higher.

A second general piece of advice is to err on the side of large sample sizes. In many studies the point estimates were economically (or perhaps one should say educationally) significant, but standard errors were too large to determine whether impacts were statistically significant. In general, when planning data collection power calculations should be done to see what sample size is needed to obtain adequate statistical precision. Third, when administering tests one should test individuals on at least two different occasions. For studies on the first of the three questions this provides baseline data before a randomized trial is implemented, while for studies on the second and third questions such data provide instrumental variables for addressing measurement error problems. Fourth, future studies should measure not only numeracy and literacy but also many other skills, such as knowledge of science, agriculture, and health care. Fifth, it is very useful to examine all three questions in the same country using comparable

tests; this was used to obtain (admittedly rough) estimates of rates of return to investments in school quality in Ghana.

There are also lessons for future researchers that are specific to the first two questions. Most refer to studies of the impact of school characteristics on cognitive skills. First, each policy intervention must be defined in detail, and the written results must provide sufficient detail to clarify exactly what was evaluated. For example, a policy of providing textbooks must be described in terms of the level and grade(s) of schooling, the subject of the textbooks, the level of difficulty of the textbooks, the ratios at which they were provided, and the extent of training provided to the teachers. Second, randomized trials must be supervised closely to determine whether the policy was carried out as planned, otherwise the result could pertain to an intervention very different from that of the intended policy. Third, extensive monitoring is needed to ensure that non-random selection and non-random attrition do not drive the results. In some cases, biases due to these problems can be avoided, or at least measured, but in others they may be so severe that the impact of the intervention cannot be identified. Fourth, the impact of the intervention should be measured for different types of students. Thus data are needed on each student's age, sex, ethnic group, parental background, household income (perhaps proxied by information on ownership of durable goods), and pre-intervention academic performance. Fifth, the intervention should be replicable on a wide scale; an example of a study for with potential problems in this respect is the evaluation of the mother training program in Turkey. Sixth, the cost of the intervention is needed to calculate the ratio of the benefit (in terms of improved test scores) to the cost. Seventh, results from randomized trials should be compared to results based on conventional methods (which can be done using the sample of control schools) to assess the extent of bias in conventional estimates. Finally, if governments are reluctant to conduct randomized trials, one should approach non-governmental organizations, which may be more amenable to participating in such trials and may use the results more quickly.

Regarding studies of the second question, I have two suggestions. First, studies of the incomes of either wage earners or the self-employed must give serious thought to correcting problems of self-selection into different occupations. Second, research on

wage workers should probably drop government workers since their wages are less likely to reflect productivity differences than those of private sector workers.

Economists' interest in human capital and the role of education in economic growth, combined with their increasingly rigorous standards for data analysis, provides an opportunity to make significant progress in understanding the causes and consequences of education outcomes, and ultimately in raising the quality of life, in developing countries. Much remains to be done, but we now have a much clearer idea of what to do. Working with education professionals, international organizations and NGO's, economists can increase the quality and quantity of research on education in developing countries. While the costs of such research will not be trivial, the cost of not doing that research, in terms of inefficient use of educational resources, are likely to be much higher.

References

Acemoglu, Daron. 1996. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation," *Quarterly Journal of Economics*. 111(3): 779-804.

Alderman, Harold, Jere Behrman, David Ross and Richard Sabot. 1996a. "Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy," *Journal of Human Resources*. 31(1):229-254.

Alderman, Harold, Jere Behrman, David Ross and Richard Sabot. 1996b. "The Returns to Endogenous Human Capital in Paksitan's Rural Wage Labour Market." *Oxford Bulletin of Economics and Statistics*. 58(1):29-55.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King and Michael Kremer. 2000. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment". Department of Economics. MIT.

Angrist, Joshua, and Victor Lavy. 1997. "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco." *Journal of Labor Economics*. 15(1): S48-76.

Angrist, Joshua, and Victor Lavy. 1999a. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement" *Quarterly Journal of Economics*. 114(2): 533-575.

Angrist, Joshua, and Victor Lavy. 1999b. "New Evidence on Classroom Computers and Pupil Learning." National Bureau of Economic Research Working Paper #7424.

Barro, Robert. 1991. "Economic Growth in a Cross-Section of Countries," *Quarterly Journal of Economics*. 106(2):407-443.

Becker, Gary. 1995. "Human Capital and Poverty Alleviation". Human Capital Development and Operations Policy Working Paper No.52. The World Bank, Washington DC.

Behrman, Jere. 1990. "The Action of Human Resources and Poverty on One Another: What We Have Yet to Learn." Living Standards Measurement Study Working Paper No.74. The World Bank. Washington, D.C.

Behrman, Jere, and Nancy Birdsall. 1983. "The Quality of Schooling: Quantity Alone is Misleading." *American Economic Review*. 73(5):928-946.

Behrman, Jere, and James Knowles. 1999. "Household Income and Child Schooling in Vietnam." *World Bank Economic Review*. 13(2):211-256.

Behrman, Jere, Mark Rosenzweig and Paul Taubman. 1994. "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment." *Journal of Political Economy*. 102(6): 1131-1174.

Benabou, Roland. 1996. "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance". *American Economic Review*. 86(3): 584-609.

Birdsall, Nancy. 1988. "Public Inputs and Child Schooling in Brazil." *Journal of Development Economics*. 18(1):67-86.

Boissiere, Maurice, John Knight and Richard Sabot. 1985. "Earnings, Schooling, Ability and Cognitive Skills." *American Economic Review*. 75(5):1016-1030.

Bowles, Samuel, and Herbert Gintis. 1976. *Schooling in Capitalist America: Education Reform and the Contradiction of Economic Life*. New York: Basic Books.

Card, David. 1999. "The Causal Effect of Education on Earnings," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics Volume 3*. North Holland.

Card, David, and Alan Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*. 100(1):1-40.

Case, Anne, and Angus Deaton. 1999. "School Inputs and Educational Outcomes in South Africa," *Quarterly Journal of Economics*. 114(3):1047-1084.

Cox, Donald, and Emmanuel Jimenez. 1991. "The Relative Effectiveness of Private and Public Schools: Evidence from Two Developing Countries." *Journal of Development Economics*. 34(1):99-121.

Davidson, Russell, and James MacKinnon. 1993. *Estimation and Inference in Econometrics*. Oxford University Press: New York.

Duflo, Esther. Forthcoming. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*.

Pritchett, Lant, and Deon Filmer. 1999. "What Education Production Functions *Really* Show: A Positive Theory of Education Expenditures." *Economics of Education Review*. 18:223-239.

Foster, Andrew, and Mark Rosenzweig. 1993. "Information, Learning and Wage Rates in Low-Income Rural Areas." *Journal of Human Resources*. 28(4):759-790.

Fuller, Bruce. 1987. "What School Factors Raise Achievement in the Third World?" *Review of Educational Research*. 57(3):255-292.

Fuller, Bruce, and Prema Clark. 1994. "Raising School Effects while Ignoring Culture? Local Conditions and the Influence of Classroom Tools, Rules and Pedagogy." *Review of Educational Research*. 64(1):119-157.

Gertler, Paul, and Paul Glewwe. 1990. "The Willingness to Pay for Education in Developing Countries: Evidence from Peru." *Journal of Public Economics*. 42(3):251-275.

Glewwe, Paul. 1996. "The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment." *Journal of Development Economics*. 51(2):267-290.

Glewwe, Paul. 1999a. *The Economics of School Quality Investments in Developing Countries*. St. Martin's Press: New York.

Glewwe, Paul. 1999b. "School Quality, Student Achievement, and Fertility in Developing Countries," in Bledsoe, C. et al. eds., *Critical Perspectives on Schooling and Fertility in the Developing World*. National Academy Press. Washington D.C.

Glewwe, Paul. 1999c. "Why Does Mother's Schooling Raise Child Health in Developing Countries? Evidence from Morocco." *Journal of Human Resources*. 34(1):124-159.

Glewwe, Paul, and Jaikishan Desai. 1999. "Child Health and Mothers' Education in Ghana," in Glewwe, P., *The Economics of School Quality Investments in Developing Countries*. Macmillan.

Glewwe, Paul, Margaret Grosh, Hanan Jacoby, and Marlaine Lockheed. 1995. "An Eclectic Approach to Estimating the Determinants of Achievement in Jamaican Primary Education." *World Bank Economic Review* 9 (2): 231-258.

Glewwe, Paul, and Hanan Jacoby. 1992. "Estimating the Determinants of Cognitive Achievement in Low-income Countries: The Case of Ghana." Living Standards Measurement Study Working Paper No.91. The World Bank. Washington, D.C.

Glewwe, Paul, and Hanan Jacoby. 1994. "Student Achievement and Schooling Choice in Low-income Countries: Evidence from Ghana." *Journal of Human Resources* 29 (3): 843-864.

Glewwe, Paul, Michael Kremer and Sylvie Moulin. 2001. "Textbooks and Test Scores: Evidence from a Randomized Evaluation in Kenya". Development Research Group. The World Bank. Washington DC.

Glewwe, Paul, Michael Kremer, Sylvie Moulin and Eric Zitzewitz. 2000. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya". NBER Working Paper No. 8018. Cambridge, MA.

Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica*. 45(1)1-22.

Hanushek, Eric. 1995. "Interpreting Recent Research on Schooling in Developing Countries." *World Bank Research Observer* 10(2):227-246.

Hanushek, Eric. 1996. "School Resources and Student Performance," in G. Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Brookings Institution.

Harbison, Ralph, and Eric Hanushek. 1992. *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. Oxford University Press.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica*, 66(5):1017-1098.

Heckman, James, Hidehiko Ichimura and Petra Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program". *Review of Economic Studies*. 64(4): 605-654.

Heckman, James, Robert Lalonde and Jeffrey Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics: Volume 3*. North Holland.

Heyneman, Stephen P., Dean T. Jamison and Xenia Montenegro. 1984. "Textbooks in the Philippines: Evaluation of the Pedagogical Impact of Nationwide Investment." *Educational Evaluation and Policy Analysis* 6(2):139-150.

Jamison, Dean, Barbara Searle, Klaus Galda and Stephen Heyneman. 1981. "Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement". Journal of Educational Psychology 73(4): 556-67.

Jimenez, Emmanuel, and Yasuyuki Sawa. 1999. "Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program." *World Bank Economic Review*. 13(3):415-441.

Jolliffe, Dean. 1998. "Skills, Schooling and Household Income in Ghana." *World Bank Economic Review*. 12(1):81-104.

Kagitcibasi, Cigdem, Diane Sunar and Sevda Bekman. Forthcoming. "Long-Term Effects of Early Intervention: Turkish Low-Income Mothers and Children." *Journal of Applied Developmental Psychology*.

King, Elizabeth, and Berk Ozler. 2000. "What's Decentralization Got to Do with Learning: Endogenous School Quality and Student Performance in Nicaragua." Development Research Group. The World Bank. Washington, DC.

Kingdon, Geeta. 1996a. "Student Achievement and Teacher Pay: A Case Study of India." STICERD Working Paper No. 74. London School of Economics and Political Science.

Kingdon, Geeta. 1996b. "The Quality and Efficiency of Private and Public Education: A Case Study of Urban India." *Oxford Bulletin of Economics and Statistics*. 58(1): 57-82.

Kremer, Michael. 1995. "Research on Schooling: What We Know and What We Don't – A Comment on Hanushek." *World Bank Research Observer*. 10(2):247-254.

Kremer, Michael, Sylvie Moulin, Robert Namunyu and David Myatt. 1997. "The Quantity-Quality Tradeoff in Education: Evidence from a Prospective Evaluation in Kenya". Unpublished.

Lockheed, Marlaine, and Adriaan Verspoor. 1991. *Improving Primary Education in Developing Countries*. Oxford University Press: New York.

Lucas, Robert. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*. 22: 3-42.

Mankiw, N. Gregory, David Romer and David Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics*. 107(2): 407-437.

Manski, Charles, and Irwin Garfinkel. 1992. *Evaluating Welfare and Training Programs*. Harvard University Press: Cambridge, MA.

McEwan, Patrick, and Martin Carnoy. 2000. "The Effectiveness and Efficiency of Private Schools in Chile's Voucher System." *Educational Evaluation and Policy Analysis*. 22(3):213-239.

Miguel, Edward, and Michael Kremer. 2000. "Child Health and Education: The Primary School De-Worming Project in Kenya." Department of Economics. University of California at Berkeley.

Mizala, Alejandra, and Pilar Romaguera. 2000. "School Performance and Choice." *Journal of Human Resources*. 35(2):392-417.

Moll, Peter. 1998. "Primary Schooling, Cognitive Skills and Wages in South Africa." *Economica*. 65:263-284.

Murnane, Richard, John Willet and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics*. 77(2):251-266.

Newman, John, Laura Rawlings and Paul Gertler. 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *World Bank Research Observer*. 9(2):181-201.

Oliver, Raylynn. 1999. "Fertility and Women's Schooling in Ghana," in Glewwe, P., *The Economics of School Quality Investments in Developing Countries*. Macmillan.

Pitt, Mark, Mark Rosenzweig, and Donna Gibbons. 1993. "The Determinants and Consequences of the Placement of Government Programs in Indonesia." *World Bank Economic Review* 7 (3): 319-348.

Powell, James. 1994. "Estimation of Semiparametric Models," in R. Engle and D. McFadden, eds., *Handbook of Econometrics: Volume IV*. North Holland.

Rosenzweig, Mark. 1995. "Why Are There Returns to Schooling?" *American Economic Review*. 85(2):153-158.

Rosenzweig, Mark., and Kenneth Wolpin. 1986. "Evaluating the Effects of Optimally Distributed Public Programs." *American Economic Review*. 76(3):470-482.

Rosenzweig, Mark., and Kenneth Wolpin. 2000. "Natural 'Natural' Experiments in Economics." *Journal of Economic Literature*. 38(4):827-874.

Rouse, Cecilia. 1998. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *Quarterly Journal of Economics*. 113(2):553-602.

Tan, Jee-Peng, Julia Lane and Gerard Lassibille. 1999. "Outcomes in Philippine Elementary Schools: An Evaluation of Four Experiments." *World Bank Economic Review*. 13(3):493-508.

Thomas, Duncan. 1999. "Fertility, Education and Resources in South Africa," in Bledsoe, C. et al. eds., *Critical Perspectives on Schooling and Fertility in the Developing World*. National Academy Press. Washington D.C.

UNDP. 1990. *Human Development Report*. United Nations Development Programme. New York.

Vijverberg, Wim. 1992. "Measuring Income from Family Enterprises with Household Surveys." *Small Business Economics*. 4:287-305.

Vijverberg, Wim. 1999. The Impact of Schooling and Cognitive Skills on Income from Non-Farm Self-Employment," in Glewwe, P., *The Economics of School Quality Investments in Developing Countries*. Macmillan.

Wolpin, Kenneth. 1977. "Education and Screening." *American Economic Review*. 67(5):949-958.

World Bank. 1998. *World Development Indicators 1998*. Washington, DC.

World Bank. 2001. *World Development Report 2000/2001: Attacking Poverty*. Washington, DC.

Yang, Dennis. 1997. "Education in Production: Measuring Labor Quality and Management." *American Journal of Agricultural Economics*. 79(3):764-772.

RECENT BULLETINS

95-5 Fahima Aziz, "Nutrition, Health and Labor Productivity Analysis of Male and Female Workers: A Test of the Efficiency Wage Hypothesis." August 1995.

95-6 Hamin H. Elbasha and Terry L. Roe, "Environment in Three Classes of Endogenous Growth Models." August 1995.

95-7 Gopinath Munisamy and Terry L. Roe, "Sources of Sectoral Growth in an Economy Wide Context: The Case of U.S. Agriculture." August 1995.

95-8 Gopinath Munisamy and Terry L. Roe, "General Equilibrium Analysis of Supply and Factor Returns in U.S. Agriculture, 1949-91." August 1995.

95-9 Douglas Gollin, "Do Taxes on Large Farms Impede Growth?: Evidence from Ghana." 1995.

96-1 X. Diao, E.H. Elbasha, T.L. Roe, and E. Yeldan, "A Dynamic CGE Model: An Application of R&D-Based Endogenous Growth Model Theory." May 1996.

96-2 Munisamy Gopinath and Terry L. Roe, "R&D Spillovers: Evidence from U.S. Food Processing, Farm Machinery and Agriculture." October 1996.

96-3 Xinshen Diao and Agapi Somwaru, "Dynamic Gains and Losses from Trade Reform: An Intertemporal General Equilibrium Model of the United States and MERCOSUR." October 1996.

96-4 Vernon Ruttan, "Sources of Technical Change: Induced Innovation, Evolutionary Theory and Path Dependence." October 1996.

97-1 Hung-Lin Tao and Thomas Stinson, "An Alternative Measure of the Human Capital Stock." April 1997.

98-1 Xinshen Diao, Terry L. Roe, and A. Erinc Yeldan, "How Fiscal (Mis)-Management May Impede Trade Reform: Lessons From an Intertemporal, Multi-Sector General Equilibrium Model for Turkey." January 1998.

98-2 Xinshen Diao, Terry Roe, and Erinc Yeldan, "Strategic Policies and Growth: An Applied Model of R&D-Driven Endogenous Growth." March 1998

98-3 Anna Maria Siti Kawuryan, "Educational Achievement and Sectoral Transition in the Indonesian Labor Force." March 1998

98-4 Vernon W. Ruttan, "Growth Economics and Development Economics: What Should Developmental Economists Learn (if anything) from the New Growth Theory?" April 1998.

98-5 Xinshen Diao and Terry Roe, "The Effect of Sequencing Trade and Water Market Reform on Interest Groups in Irrigated Agriculture: An Intertemporal Economy-Wide Analysis of the Moroccan Case." November 1998.

98-6 Willis Peterson, "Monetary Instability and Economic Growth." November 1998.

99-1 Pei-Chien Lin and Terry Roe, "Growth Theory and Accounting for Growth of the Taiwanese Economy." July 1999.

99-2 H. Colby, X. Diao, and A. Somwaru, "Sources of Growth and Supply Response: A Cross-Commodity Analysis of China's Grain Sector." July 1999.

00-1 Beatriz Gaitan and Ferdinand Pavel, "Is 'Getting the Prices right' Always Right? How Trade Liberalization Can Fail." March 2000.

00-2 Vernon W. Ruttan, "Imperialism and Competition in Anthropology, Sociology, Political Science and Economics: A Perspective from Development Economics." March 2000.

01-1 Xinshen Diao, Agapi Somwaru and Terry Roe, "A Global Analysis of Agricultural Trade Reform in WTO Member Countries." January 2001.

01-2 Hamid Mohtadi and Terry Roe, "Democracy, Rent Seeking, Public Spending and Growth." June 2001.

BEST COPY AVAILABLE

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Urban Education
Box 40, Teachers College
Columbia University
525 W. 120th Street, Main Hall 303
New York, NY 10027

Tel: 212-678-3433 / 800-601-4868
Fax: 212-678-4012

http://eric-web.tc.columbia.edu

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)